SLAC-PUB-XXXXX

Learning How to Count: A High Multiplicity Search for the LHC

Sonia El Hedri,[1] Anson Hook,[2] Martin Jankowiak,[3] and Jay G. Wacker[1]

[1] *SLAC, Stanford University, Menlo Park, CA 94025 USA*

[2] *Institute for Advanced Studies, Princeton University, Princeton, NJ 08544 USA*

[3] *Institut für Theoretische Physik, Universität Heidelberg, Germany*

Abstract

We introduce a search technique that is sensitive to a broad class of signals with large final state multiplicities. Events are clustered into large radius jets and jet substructure techniques are used to count the number of subjets within each jet. The search consists of a cut on the total number of subjets in the event as well as the summed jet mass and missing energy. Two different techniques for counting subjets are described and expected sensitivities are presented for eight benchmark signals. These signals exhibit diverse phenomenology, including 2-step cascade decays, direct three body decays, and multi-top final states. We find improved sensitivity to these signals as compared to previous high multiplicity searches as well as a reduced reliance on missing energy requirements. One benefit of this approach is that it allows for natural data driven estimates of the QCD background.

Contents

1. INTRODUCTION

The search for physics beyond the Standard Model is a central focus of LHC research. The motivations for extensions of the Standard Model are multi-faceted, spanning such diverse physics topics as the identity of dark matter, the radiative stability of the weak scale, the unification of forces, and the origin of the baryon asymmetry of the Universe. Apart from these specific theory inputs, which suggest certain classes of models, there is the generic goal of thoroughly probing the weak scale for new physics—whatever that might be. In order to cover the vast range of possibilities for new physics that open up when the input from theory is loosened, it is imperative for the LHC to carry out an extensive experimental program that is sensitive to the widest possible range of new physics signatures.

Since it is not possible to perform model independent searches for new physics—doing so is limited by both theoretical and experimental systematic uncertainties—it is necessary to design searches for new physics that are targeted at specific experimental signatures. Typically such searches are based on exploiting a single key handle that significantly reduces the dominant Standard Model backgrounds, namely those originating from QCD jet production. For instance, it is common to require hard electroweak particles such as leptons or photons or (large amounts of) missing energy. Requiring b-tagged jets, massive jet resonances or extremely high energy events can provide alternative avenues for parametrically reducing the QCD background. These considerations exist both because of triggering requirements necessary for permanently storing events to tape and because realistic systematic uncertainties in any case limit the degree to which increased integrated luminosity leads to increased sensitivity.

The past several years have seen increasing attention being paid to high multiplicities (i.e. requiring that events have more than ~ 6 final state jets) as a way of parametrically reducing QCD contributions to searches for new physics. Historically this approach was motivated by searches for black holes at the LHC (see e.g. ref. [1]), but more recently high multiplicity searches have been advocated as an effective technique for helping to probe scenarios with natural supersymmetry as well as those with baryonic R-parity violating supersymmetry [2–7]. High multiplicity final states also arise in theories of strong dynamics where new colored objects can produce four to eight top quarks through the production of "coloron" vector resonances or colored technipions [8, 9]. Finally, some models introduced to explain the magnitude of the $t\bar{t}$ asymmetry measured at the Tevatron also predict large final state multiplicites [10].

One of the challenges of using high multiplicity as a handle for reducing QCD backgrounds is that the background rate is intrinsically difficult to calculate. The current state of the art for tree-level jet production is $2 \to 7$, with $2 \to 6$ being the most that is typically feasible. Significantly, these tree-level calculations have unquantified uncertainties in their rates and distributions. One of the computational challenges is that high multiplicity final states have enormous configuration spaces; for instance the 10 jet final state has 28 dimensions, with the consequence that is unfeasible to densely populate the configuration space with Monte Carlo events. Many of the configuration space variables, such as the angular separations between jets, ΔR_{ij}, have not been studied in depth and historically have been unreliably calculated with tree-level Monte Carlo.

One way that high multiplicity backgrounds are estimated is by extrapolating from lower multiplicities. There are well-known approximate empirical scaling relations connecting the N jet production rate to the $N+1$ rate. There has been some progress in deriving these from first principles (see e.g. ref. [11, 12]); nevertheless, this approach comes with large uncertainties in the rates that will cause these searches to quickly become systematically limited. Additionally, for large N the p_T spectrum for the Nth jet becomes increasingly soft and harder to measure accurately, leading to additional uncertainties.

Recently alternative approaches to gaining sensitivity to high multiplicity final states have been developed. In effect these proposals factorize the problem: instead of directly counting jets, the event is clustered into a fixed number of large radius jets ($N = 4$, 5, or

6) whose substructure is then further scrutinized. Because the jet radius is large, these N "fat" jets will incorporate most of the radiation in the central region of the detector. The particles that would have formed multiple small radius jets in the traditional approach are clustered together into large radius jets. These fat jets will automatically have substructure, some of which will appear to originate from hard 1 → 2 parton shower splittings. While such splittings certainly occur in QCD, they occur relatively rarely with the result that requiring multiple fat jets to have multiple hard splittings helps to separate signal events with high multiplicities from the dominant low multiplicity QCD background.

The fat jet approach has an additional benefit in that it may be better suited to getting a good handle on systematic uncertainties in the QCD backgrounds. Large radius jets, particularly well separated ones, have properties that are relatively independent from one another, since their dynamics are largely driven by the parton shower, which is local in nature. For instance, the mass of one fat jet is not strongly correlated with the masses of other fat jets in the event. By exploiting the approximate independence of jets, QCD backgrounds can be estimated with a data driven analysis. Effectively the large configuration space we started out with has been factorized into a much smaller fat jet configuration space tied to N (approximately identical) configuration spaces encoding the fat jets' substructure. This factorization lends itself to the measurement of appropriate jet templates, which can then be combined with fat jet distributions to arrive at background estimates. This approach is of obvious practical importance to experimental searches, but it is also useful for theoretical calculations of the backgrounds, since searching for obvservables that reduce QCD backgrounds by six to ten orders of magnitude while acquiring enough statistics in Monte Carlo can be challenging to prohibitive.

Specifically, this study builds on a recent paper [13] that proposed M_J, the sum over fat jet masses, as an effective observable for separating high multiplicity signals from Standard Model backgrounds. Jet mass is the simplest jet substructure observable, but it is also one of the coarsest. This article advocates a more refined use of jet substructure to probe high multiplicity events, in particular by counting the number of subjets inside each fat jet. This may seem similar to clustering events into small radius jets and counting the resulting number of jets, but the potential to systematically estimate QCD backgrounds with a data driven approach distinguishes it from the traditional approach.

This article is organized as follows. Sec. 2 presents the two subjet counting techniques introduced in this paper. Sec. 3 describes how the backgrounds were generated in Monte Carlo. Sec. 4 describes how the backgrounds were validated against ATLAS data and how the QCD backgrounds should be amenable to data driven estimates. Sec. 5 introduces eight benchmark signals and presents the expected sensitivity of our search strategy. A comparison to previous high multiplicity searches is also made. We conclude in Sec. 6 with some general discussion.

2. COUNTING SUBJETS

This section describes the two subjet counting techniques implemented in this study, $n_{\mathrm{k_T}}$ and n_{CA}. The former is a straightforward application of the exclusive k_T algorithm [14]. The latter counts subjets by recursively inspecting the structure of the Cambridge/Aachen [15] clustering tree of the fat jet. Both are implemented using `FastJet` 3 [16]. As we will see, searches incorporating these methods yield improved sensitivity to the high multiplicity signals considered in Sec. 5. The two algorithms result in very similar expected limits, with the difference that n_{CA} does better than $n_{\mathrm{k_T}}$ in cases where the QCD backgrounds are more important.

2.1. Wide radius jets

Wide radius jets are now a standard technique in high energy physics searches. The basic structure of our high multiplicity search is built around such fat jets and is common to both subjet counting techniques. First the event is clustered into fat jets with $R_0 = 1.2$ using the anti-k_T jet algorithm [17]. Next the fat jets are trimmed [18] with the parameters $r_{\mathrm{cut}} = 0.3$ and $f_{\mathrm{cut}} = 0.05$. While fat jets are particularly sensitive to pile-up, making some sort of jet grooming necessary, it has been shown that jet substructure observables such as jet mass and N-subjettiness [19] have a significantly reduced sensitivity to pile-up effects with this choice of parameters [20]. Next the leading fat jet is required to have $p_T > 100\,\mathrm{GeV}$, while subleading fat jets are required to have $p_T > 50\,\mathrm{GeV}$. Only those events with four or more such fat jets are considered. Then the subjet count n_i of each of the four leading fat jets is calculated using one of the two algorithms described below. Finally cuts are made on the observables

$$M_J \equiv \sum_{i=1}^{4} m_i \qquad\qquad N_J \equiv \sum_{i=1}^{4} n_i \tag{1}$$

and the missing transverse energy ($\not\!\!E_T$).

2.2. Counting with k_T

The exclusive k_T algorithm is defined via two metrics, d_{ij} and d_{iB}, and a dimensionful resolution parameter d_{cut} [21]. The jet-jet metric d_{ij} and the jet-beam metric d_i are defined as:

$$d_{ij} = \min\left[p_{Ti}^2, p_{Tj}^2\right] \Delta R_{ij}^2 \qquad\qquad d_i = p_{Ti}^2 \tag{2}$$

Here, p_{Ti} is the transverse momentum of protojet i and $\Delta R_{ij}^2 \equiv \Delta\eta_{ij}^2 + \Delta\phi_{ij}^2$. The exclusive mode of the algorithm proceeds by sequentially clustering pairs of protojets, stopping once all the d_{ij} and d_i are above d_{cut}. When the smallest metric is d_{ij}, i and j are combined. When the smallest metric is d_i, protojet i is set aside as a beam jet. The total number of subjets (in the exclusive sense) is then given by the number of protojets that are not beam jets and that remain once the clustering step has terminated.[1] For our particular application we define $n_{\mathrm{k_T}}$ by taking

$$\sqrt{d_{\mathrm{cut}}} = f_{\mathrm{k_T}} p_{TJ}, \tag{3}$$

where p_{TJ} is the total transverse momentum of the fat jet and $f_{\mathrm{k_T}}$ is a dimensionless parameter that we take to be given by

$$f_{\mathrm{k_T}} = 0.04$$

[1] The beam jets, in the rare case they appear in the reclustering of our fat jets, are soft and at the periphery of the fat jet so the fact that they are discarded is just as we should like.

throughout. This value of $f_{\mathrm{k_T}}$ leads to good separation between signal and background, although for the range of signals considered the separation depends only weakly on the particular value used. Then for a given fat jet $n_{\mathrm{k_T}}$ is taken to be the number of subjets identified by exclusive k_T with

$$p_T \geq p_{T\,\mathrm{cut}} = 40\,\mathrm{GeV}.$$

The dependence on the parameter p_{Tcut} is rather weak for the massive fat jets of interest, which contain softer subjets only infrequently.

A significant advantage of this definition of $n_{\mathrm{k_T}}$ is that a typical QCD jet will, due to its asymmetric energy sharing (a hard core surrounded by soft radiation), have a small number of subjets since much of the soft radiation will be clustered with the core (see Fig. 1). This is in contrast to a naive application of Cambridge-Aachen for reclustering, which can yield a large number of subjets even for a single-pronged QCD jet.

2.3. Counting with Cambridge-Aachen

The $n_{\mathrm{k_T}}$ algorithm introduced in Sec. 2.2 was entirely generic in its motivation and approach. The present method, denoted by n_{CA}, aims instead to count the number of hard partons *consistent with the decay of a massive particle.* The n_{CA} algorithm is explicitly constructed to:

- identify massive substructure; and
- 'undercount' the number of subjets within a given fat jet if the energy sharing among the subjets is very asymmetric.

The latter requirement is made because asymmetric sharing of energy between subjets is a telltale sign of subjets generated via the parton shower. The method is in the spirit of the various substructure algorithms that have emerged since the introduction of the BDRS procedure [22] and which make use of the information encoded in the clustering tree of the jet. In particular, it is closely related to an intermediate step in the HEPTopTagger [23, 24].

We determine the number of subjets by unclustering the fat jet down to the mass scale m_{cut}, throwing out subjets with an asymmetric energy sharing as defined by y_{cut}. The number of identified subjets that then pass an additional p_T cut yields n_{CA}. In detail, the method is defined as follows:

1. Cluster a given fat jet using the Cambridge/Aachen algorithm.
2. To define n_{CA} inspect the clustering tree of the fat jet via the following recursive procedure.
3. Uncluster j into j_1 and j_2 with $p_{\mathrm{T1}} > p_{\mathrm{T2}}$.
4. If $m_j < m_{\mathrm{cut}}$ or $dR(j_1, j_2) < R_{\mathrm{min}}$ consider j a subjet and exit the recursion.
5. If $p_{\mathrm{T2}} < y_{\mathrm{cut}} \cdot (p_{\mathrm{T1}} + p_{\mathrm{T2}})$ throw out j_2.
6. Continue the recursion on j_1 and (if it is retained) j_2.

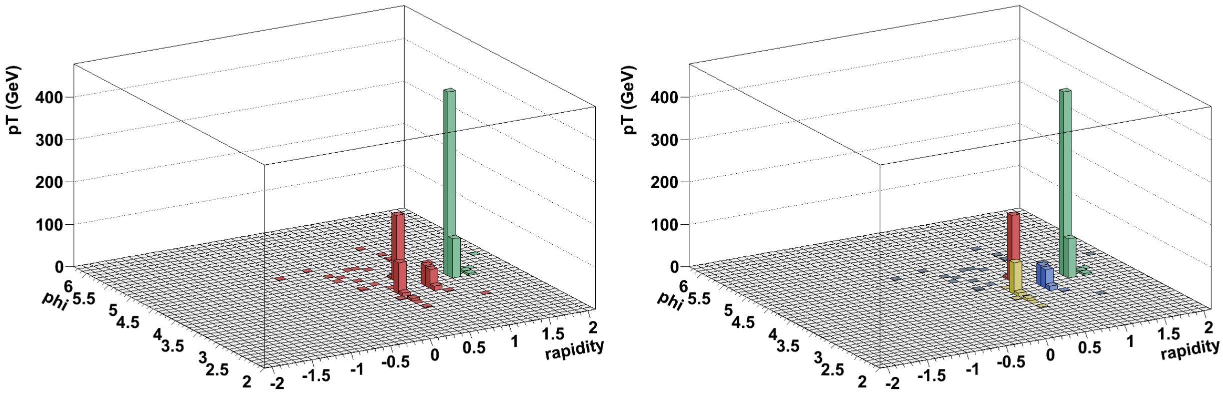

FIG. 1: Example of a fat (and very massive) QCD jet with $p_T = 910$ GeV, $m = 360$ GeV and with its two n_{k_T} subjets (left) and its four $n_{\rm CA}$ subjets (right) indicated. Note that this jet has not been trimmed to better illustrate the different treatment of soft radiation in $n_{\rm CA}$ and n_{k_T} (the dark gray cells on the right do not belong to any identified subjets).

7. When the recursion is complete, count the number of identified subjets with $p_{\rm T} > p_{\rm Tcut}$; this number is $n_{\rm CA}$.

So, for example, an idealized two-pronged jet initiated by the hadronic decay of an energetic Z boson would yield (supposing that the decay angle is such as to yield a roughly symmetric energy sharing) a count $n_{\rm CA} = 1$ for $m_{\rm cut} > m_Z$ and $n_{\rm CA} = 2$ for $m_{\rm cut} < m_Z$.

Throughout this study we use the following parameters:

$$m_{\rm cut} = 30\,{\rm GeV}, \quad y_{\rm cut} = 0.10, \quad R_{\rm min} = 0.15, \quad p_{\rm Tcut} = 30\,{\rm GeV}. \tag{4}$$

These values lead to good separation between signal and background, although for the range of signals considered the separation provided by $n_{\rm CA}$ depends only weakly on the particular values used.

2.4. Comparison of n_{k_T} and $n_{\bf CA}$

This section has introduced two distinct subjet counting techniques, and it is interesting to ask how they are related. A detailed comparison between the two algorithms is complicated by the fact that each is defined by several parameters. For simplicity we restrict ourselves to the parameter choices made above. Qualitatively, the two algorithms have a number of similar features.

On a jet-by-jet basis, there are strong correlations between n_{k_T} and $n_{\rm CA}$, with $n_{\rm CA}$ typically yielding more subjets than n_{k_T}. Fig. 1 illustrates a pronounced example of the tendency for $n_{\rm CA}$ to identify more subjets. For a given ensemble of jets, it is useful to define the normalized distribution $P(n)$, which is the fraction of jets with n subjets. The left panel of Fig. 2 shows $P(n)$ for both algorithms for a sample of leading QCD jets generated by `MadGraph` and with $p_T \geq 100\,{\rm GeV}$. Also illustrated is $P(n)$ for a sample[2] of leading jets drawn from signal events where pair produced gluinos decay via $\tilde{g} \to t\bar{t}\chi$ ($m_{\tilde{g}} = 600\,{\rm GeV}$ and $m_\chi = 60\,{\rm GeV}$).

[2] The `MadGraph` and signal-like samples, which are used throughout this section, are described in more

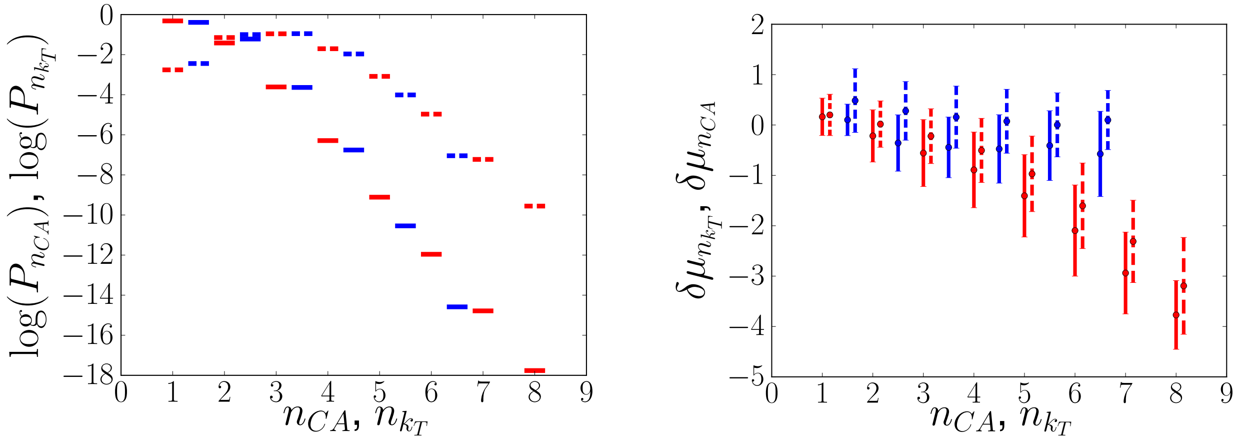

FIG. 2: *Left:* The subjet distributions $\log_{10} P_{n_{k_T}}(n_{k_T})$ and $\log_{10} P_{n_{\rm CA}}(n_{\rm CA})$ in blue and red, respectively. *Right:* The blue and red distributions show $\delta\mu_{n_{\rm CA}}(n_{k_T})$ and $\delta\mu_{n_{k_T}}(n_{\rm CA})$, respectively, which are defined in Eq. 7. The error bars on $\delta\mu_{n_{\rm CA}}(n_{k_T})$ and $\delta\mu_{n_{k_T}}(n_{\rm CA})$ correspond to the standard deviations $\sigma_{n_{\rm CA}}$ and $\sigma_{n_{k_T}}$. For both panels the solid lines correspond to a sample of leading QCD jets generated by `MadGraph` and with $p_T \geq 100\,$GeV, while dashed lines correspond to a signal-like sample described in the text.

To study how n_{k_T} and $n_{\rm CA}$ are correlated, it is useful to introduce the joint distribution

$$P(n_{k_T}, n_{\rm CA}) \qquad \text{with} \qquad \sum_{n_{k_T}, n_{\rm CA}} P(n_{k_T}, n_{\rm CA}) = 1 \tag{5}$$

From the joint distribution one can define the mean of $n_{\rm CA}$ as a function n_{k_T} as well as the mean of n_{k_T} as a function of $n_{\rm CA}$:

$$\mu_{n_{\rm CA}}(n_{k_T}) = \sum_{n_{\rm CA}} n_{\rm CA} P(n_{k_T}, n_{\rm CA}) \quad \text{and} \quad \mu_{n_{k_T}}(n_{\rm CA}) = \sum_{n_{k_T}} n_{k_T} P(n_{k_T}, n_{\rm CA}) \tag{6}$$

From this one can then define the quantities

$$\delta\mu_{n_{\rm CA}}(n_{k_T}) = \mu_{n_{\rm CA}} - n_{k_T} \qquad \text{and} \qquad \delta\mu_{n_{k_T}}(n_{\rm CA}) = \mu_{n_{k_T}} - n_{\rm CA} \tag{7}$$

which are shown in the right panel of Fig. 2. It can be seen that that n_{k_T} and $n_{\rm CA}$ track one another pretty closely for small n, but that for larger numbers of subjets $n_{\rm CA}$ tends to pull further and further ahead of n_{k_T}. The correlation between n_{k_T} and $n_{\rm CA}$ is somewhat tighter in the signal-like sample than in the QCD sample.

For fixed n_{k_T} the distribution $P(n_{k_T}, n_{\rm CA})$ is a function of $n_{\rm CA}$ with standard deviation $\sigma_{n_{\rm CA}}(n_{k_T})$. The standard deviation $\sigma_{n_{\rm CA}}(n_{k_T})$ is a steadily rising function of n_{k_T}, as illustrated by the error bars in the right panel of Fig. 2. For the QCD sample it grows from $\sigma_{n_{\rm CA}}(1) \simeq 0.3$ to $\sigma_{n_{\rm CA}}(6) \simeq 1.0$, indicating that for small n the two algorithms identify very similar numbers of subjets, but that as the amount of substructure grows there is less

detail in Sec. 3.2.1 and 5.1, respectively. The `MadGraph` samples are used because of the superior statistics available.

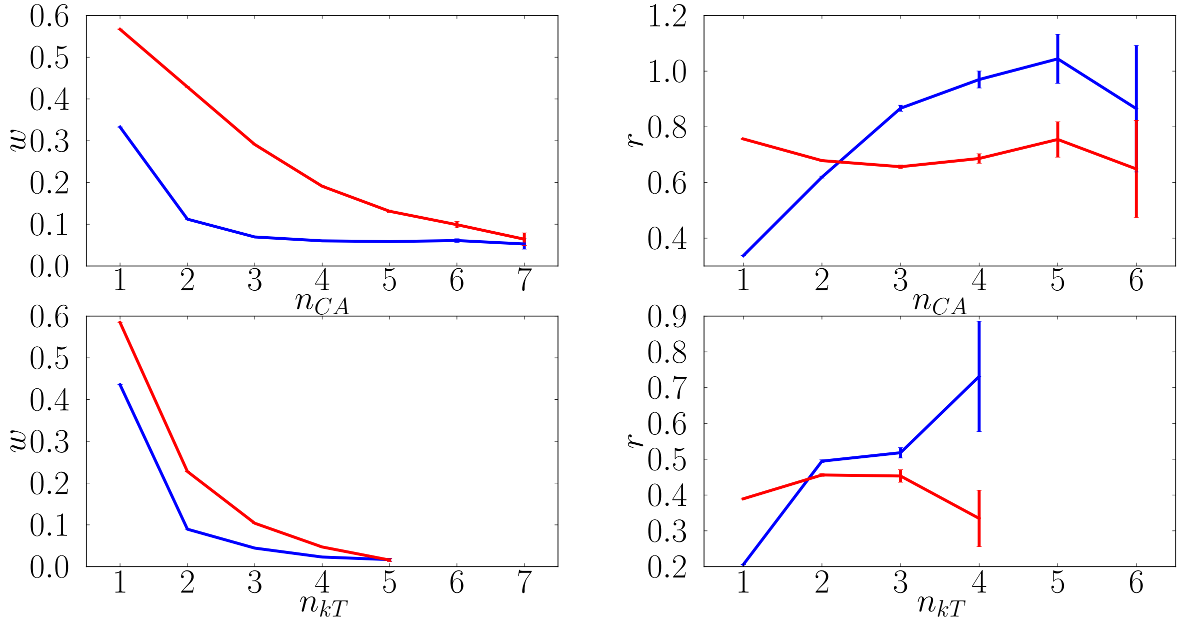

FIG. 3: Scaling patterns for the subjet distributions of leading QCD fat jets generated by `MadGraph`. *Left*: The ratio $w(n)$ for n_{k_T} (bottom) and $n_{\rm CA}$ (top) for jets with $p_T \geq 100\,\text{GeV}$ (blue) and $p_T \geq 500\,\text{GeV}$ (red). "Staircase scaling" corresponds to a flat $w(n)$. *Right*: The ratio $r(n)$ for n_{k_T} (bottom) and $n_{\rm CA}$ (top) for jets with $p_T \geq 100\,\text{GeV}$ (blue) and $p_T \geq 500\,\text{GeV}$ (red). "Double staircase scaling" corresponds to a flat $r(n)$ and appears to be emerging in the high p_T regime.

agreement between the two algorithms. Similarly the standard deviation $\sigma_{n_{k_T}}(n_{\rm CA})$ grows approximately linearly from $\sigma_{n_{k_T}}(1) \simeq 0.4$ to $\sigma_{n_{k_T}}(8) \simeq 0.9$. Note that this dispersion is logically distinct from the divergence in the mean seen in $\delta\mu_{n_{\rm CA}}(n_{k_T})$.

Note that both $n_{\rm CA}$ and n_{k_T} are peaked at 3 for the leading jet of the signal (see Fig. 2). This is as expected, since this signal contains up to 12 final state quarks, with the result that, if the leading four fat jets capture all of the decay products, an average of 3 subjets per fat are expected.

Interestingly, the subjet distributions for QCD fat jets are not governed by approximate "staircase scaling," as one might have expected. This is a scaling pattern defined by the condition that, if we define the ratio

$$w(n) \equiv \frac{P(n+1)}{P(n)} \tag{8}$$

then $w(n)$ is a constant independent of n. Instead, a significantly steeper distribution is seen. The variable

$$r(n) \equiv \frac{w(n+1)}{w(n)} = \frac{P(n)P(n+2)}{P(n+1)^2} \tag{9}$$

is useful for characterizing deviations from constant w, with staircase scaling corresponding to $r = 1$. For the case of QCD fat jets the ratios $w(n)$ and $r(n)$ for both n_{k_T} and $n_{\rm CA}$ are illustrated in Fig. 3. Since there are intrinsic energy scales in both n_{k_T} and $n_{\rm CA}$, low p_T jets are sensitive to these scales. For $p_T \geq 100\,\text{GeV}$, the $w(n)$ distribution drops rapidly before asymptoting to a constant of $w_{n_{\rm CA}} \simeq 0.05$ and $w_{n_{k_T}} \simeq 0.03$, indicating staircase scaling with a very hard spectrum. Requiring higher p_T jets makes the influence of the intrinsic scales in the counting algorithms less relevant. If the minimum p_T of the jet is raised to 500 GeV, the scales inside the subjet counting algorithm will play a much smaller role. Even for these high p_T jets staircase scaling is *not* observed; instead $w(n)$ appears to have a staircase behavior and $r(n)$ is approximately constant with $r_{n_{\rm CA}} \simeq 0.5$ and $r_{n_{k_T}} \simeq 0.4$. This can be called "double staircase scaling" and indicates a distribution of subjets that scales like

$$P(n) \sim r^{n^2/2}.$$

This is in contrast to the more traditional staircase scaling that gives

$$P(n) \sim w^n$$

demonstrating that subjet counting is not related in a straightforward manner to jet counting. Deviations from staircase scaling is well known and have even been explained in [12]; however, the deviations from staircase scaling noted here do not appear to be "Poisson" which predicts that r should asymptote to unity which does not appear to happen.

3. MONTE CARLO CALCULATIONS

This article studies the use of the total number of subjets in an event as a way to separate new physics scenarios that produce many final state quarks and gluons from QCD and electroweak-scale backgrounds. The ultimate goal is to reduce the reliance upon missing transverse energy ($\not\!\!E_T$) and lepton requirements so as not to veto on signals that have neither, while still obtaining relatively low background search regions. Since $\not\!\!E_T$ and leptons are the standard handles used to reduce QCD backgrounds, it is particularly important to have reliable estimates of multijet production rates and differential distributions. Of course, the same holds true for the non-QCD backgrounds. Consequently this section and the next play an important role in everything that follows. Throughout, the calculations are performed at a center of mass energy of $\sqrt{s} = 8\,\text{TeV}$.

The rest of this section is organized as follows. Sec. 3.1 describes the calculation of non-QCD backgrounds such as V+jets and $t\bar{t}$+jets. Sec. 3.2 describes the two approaches used to generate QCD Monte Carlo events. Sec. 3.3 describes how detector effects and jet clustering are implemented. These latter two subsections are complemented by the discussion in Sec. 4, which focuses on data driven methods.

3.1. Non-QCD Backgrounds

The dominant Standard Model backgrounds are QCD, V+jets and $t\bar{t}$+jets. Since, however, any backgrounds where there is an intrinsic mass scale are potentially important, the leading subdominant backgrounds were also computed for completeness, see Table 1. The non-QCD backgrounds used for our analyses were generated using `MadGraph 4.5.1` [25–27] and showered and hadronized using `PYTHIA 6.4` [28] . The five-flavor MLM matching scheme with a shower-$k_\perp$ scheme was used to account for the extra radiation [29].

It is the high p_T tails of these backgrounds that will make the dominant contribution to signal regions. Since some of these backgrounds have quite large cross sections, it is not

Process	Order	p_T range (GeV)	Cross Section (fb)	Events	Event Weight
$V + n_3 j$	$\mathcal{O}(\alpha_w \alpha_s^3)$	0 - 100	8,586,000	2.5×10^6	103.0
	$\mathcal{O}(\alpha_w \alpha_s^3)$	100- 200	949,000	2.5×10^6	11.4
	$\mathcal{O}(\alpha_w \alpha_s^3)$	200- 300	72,400	2.5×10^6	0.87
	$\mathcal{O}(\alpha_w \alpha_s^3)$	300+	15,200	2.5×10^6	0.18
$t\bar{t} + n_2 j$	$\mathcal{O}(\alpha_s^4), \mathcal{O}(\alpha_w^2 \alpha_s^2)$	0-150	94,300	3.5×10^5	8.08
	$\mathcal{O}(\alpha_s^4), \mathcal{O}(\alpha_w^2 \alpha_s^2)$	150-300	33,900	3.5×10^5	2.90
	$\mathcal{O}(\alpha_s^4), \mathcal{O}(\alpha_w^2 \alpha_s^2)$	300+	4,440	3.5×10^6	0.38
$VV' + n_2 j$ *	$\mathcal{O}(\alpha_w^2 \alpha_s^2)$	0+	51,500	5.4×10^5	2.86
$Vt + n_2 j$ *	$\mathcal{O}(\alpha_w \alpha_s^3)$	0-200	13,870	1.8×10^5	2.31
	$\mathcal{O}(\alpha_w \alpha_s^3)$	200-300	984	1.8×10^5	0.16
	$\mathcal{O}(\alpha_w \alpha_s^3)$	300+	319	1.8×10^5	0.053
$t + n_3 j$	$\mathcal{O}(\alpha_w^2 \alpha_s^2)$	0-200	11,800	2.6×10^5	1.32
	$\mathcal{O}(\alpha_w^2 \alpha_s^2)$	200-300	1,430	2.6×10^5	0.16
	$\mathcal{O}(\alpha_w^2 \alpha_s^2)$	300+	355	2.6×10^5	0.040
VH	$\mathcal{O}(\alpha_w^2)$	0+	975	3.0×10^4	0.98
$t\bar{t}V + n_1 j$	$\mathcal{O}(\alpha_w \alpha_s^3)$	0+	310	3.0×10^4	0.31
$t\bar{t}H + n_1 j$	$\mathcal{O}(\alpha_w \alpha_s^3)$	0+	130	3.0×10^4	0.13
$t\bar{t}t\bar{t}$	$\mathcal{O}(\alpha_s^4)$	0+	0.8	1.0×10^4	0.00024

TABLE 1: The non-QCD backgrounds used in this analysis. The subscript i in n_i indicates the highest jet multiplicity considered in the matched sample. Thus $t\bar{t} + n_2 j$ is $t\bar{t} + 0j$, $t\bar{t} + 1j$, $t\bar{t} + 2^+ j$, where the last jet multiplicity is an inclusive process that can include higher jet multiplicities generated through the parton shower. The p_T slicing is with respect to the leading massive object. The two samples marked with a $*$ denote that resonant top production is excluded to avoid double counting. The last column indicates the ratio between the expected number of events at $30\,\text{fb}^{-1}$ and the number of Monte Carlo events generated.

feasible to generate $30\,\text{fb}^{-1}$ worth of Monte Carlo. Instead, the backgrounds are generated in different p_T bins of the heavy particle, $p_{T\,\text{heavy}}$, where a heavy particle is any one of t, $W^\pm$, Z^0, or H^0.[3] This ensures that the regions of phase space that result in the largest contamination of the signal region are computed with sizeable significance. Another important consideration is that events with small $p_{T\,\text{heavy}}$ can still have large H_T as a consequence of high jet activity. Since these backgrounds are potentially important for regions of phase space with sizeable $\not\!\!E_T$, it is important for them to be computed with sufficient significance. However, if $p_{T\,\text{heavy}}$ is small, then so is the intrinsic $\not\!\!E_T$, with the consequence that the high H_T/low $p_{T\,\text{heavy}}$ region of phase space is not important for the signal region because it is subdominant to the QCD contribution. In practice, the large event weights of the low $p_{T\,\text{heavy}}$ regions do not limit the statistical accuracy of the background estimate.

The resulting backgrounds are shown in Table 1, where the subscript i in n_i indicates the highest jet multiplicity considered in the matched sample. The different $p_{T\,\text{heavy}}$ bins

[3] When there are two or more heavy particles in the event, e.g. $t\bar{t}$, $p_{T\,\text{heavy}}$ denotes the larger p_T.

are listed in the third column. The five flavor matching scheme introduces one additional complication for diboson and single top production. This arises because there can be resonant top production in these two channels that has already been included in ordinary $t\bar{t}$ production. In order to exclude this double counting, a requirement of $|m_t - m_{bW}| > 15\Gamma_t$ is imposed.

The two most important non-QCD backgrounds for our search are V+jets and $t\bar{t}$+jets. These backgrounds not only have large cross sections but are also jet rich and result in a reasonable amount of $\not{E}_T$. Backgrounds like $t\bar{t}t\bar{t}$, $t\bar{t}V$ and $t\bar{t}H$, which have jet multiplicities and $\not{E}_T$ comparable to that of some of the benchmark signals we study, have small cross sections and make a negligible contribution to the total background (although we include them for completeness).

3.2. QCD

Several techniques are used to calculate the QCD contribution to signal regions. Sec. 3.2.1 describes a calculation of the QCD background using an MLM-matching scheme implemented in MadGraph and PYTHIA using unweighted events with up to four partons matched. This is a relatively low multiplicity method of generating backgrounds and relies heavily on the parton shower to generate high multiplicities; nevertheless, it is a standard calculational method that makes it easy to to get good Monte Carlo statistics over the entire signal region. Sec. 3.2.2 describes a calculation of the QCD background using a CKKW-matching scheme implemented in SHERPA using weighted events with up to six partons matched. This method allows significantly higher multiplicities to be generated and samples the high energy and high multiplicity tails with weighted events. The use of weighted events tends to hurt convergence and gives relatively poor Monte Carlo statistics.

3.2.1. MadGraph+PYTHIA

One set of QCD backgrounds was generated with MadGraph 4.5.1 [25–27] at $\mathcal{O}(\alpha_s^4)$ and showered with PYTHIA 6.4 [28] using MLM matching [29]. The events were generated in multiple exclusive samples varying both the matrix-element parton multiplicity from $n_j = 2$ to $n_j = 4$ and the scalar transverse energy H_T using the 5-flavor matching scheme [30]. It is not practical to generate multiplicities higher than $n_j = 4$ in MadGraph due to computational limitations. Since the event selection for our search strategy requires $n_j \geq 4$, all jet substructure will be modeled by the parton shower. This is clearly insufficient for gaining much confidence in the resulting background estimates. Nevertheless, the MadGraph events will serve as a useful crosscheck in what follows. In addition, the high statistics available from MadGraph will be very useful in creating missing energy templates in Sec. 4.1.

3.2.2. SHERPA

The second event generator used to generate QCD backgrounds is SHERPA 1.4.0 [31–35]. SHERPA uses CKKW matching [36] and is capable of generating up to $n_j = 6$ at the matrix element level. SHERPA can therefore generate more hard substructure using matrix elements without relying on the parton shower. Consequently we will primarily be relying on SHERPA for our QCD background estimates. This sample was generated and fully described in [47][4].

One of the drawbacks of SHERPA is that it is not straightforward to generate separate samples binned by H_T. Instead weighted Monte Carlo events can be generated. The main problem with relying on weighted Monte Carlo is that it becomes more difficult to obtain convergence in the signal region. One frequently observes that a single (or handful of) high weight event(s) can make large contributions to the tails of distributions, with the consequence that statistical uncertainties in the tails become large.

Two separate weighting methods were used in generating the QCD backgrounds. The first is the default weighting procedure in SHERPA. The second skews the weights towards higher multiplicities, with

$$w(n_j) = 4^{n_j - 2} \tag{10}$$

for $2 \leq n_j \leq 6$. Thus, 256 times as many $n_j = 6$ events are generated as compared to $n_j = 2$ events. This allowed for the generation of relatively more high multiplicity events so as to better fill out the tails of the distributions. Together these two weighting methods resulted in a total of 4.8×10^6 events passing our basic fat jet requirements (see Sec. 2.1).

3.2.3. Comparison of MadGraph *and* SHERPA

In Figs. 4 and 5 the SHERPA results are compared to MadGraph+PYTHIA. We see that there is generally good agreement between the two. Even the subjet count N_{CA} shows good agreement up to $N_{\text{CA}} = 8$, a regime in which both generators (especially MadGraph) are relying on the parton shower to generate substructure. The biggest differences appear in the tails of the distributions. As discussed above, the presence of high weight events in SHERPA can lead to poor convergence. Whenever there is a large disagreement between the two generators in these figures, it seems to be largely driven by this effect (c.f. the large statistical uncertainties in the regions of largest disagreement).

The disagreement in the tails of the distributions between SHERPA and MadGraph + PYTHIA deserves further comment. While the naive expectation is that SHERPA should provide a better description in the high H_T, M_J and N_J regime, there are significant statistical uncertainties in the SHERPA sample arising from slow convergence of the weighted Monte Carlo calculations. The features of the distributions also appear to be slightly pathological in their behavior, appearing as inflection points in regions that one would expect to be relatively smooth. In the case of N_{CA}, some of the bins are not even monotonically decreasing. This article will rely on the SHERPA calculation for its background estimates; fortunately the design of the search regions in Sec. 5 will not be heavily influenced by these features. In Sec. 4, a data driven approach to calculating these backgrounds is presented, and the disagreement between the data driven approach and the straight Monte Carlo calculation will be, at least in part, related to these same features.

[4] We thank Tim Cohen, Mariangela Lisanti, Eder Izaguirre, and Tim Lou for letting us use this Monte Carlo sample.

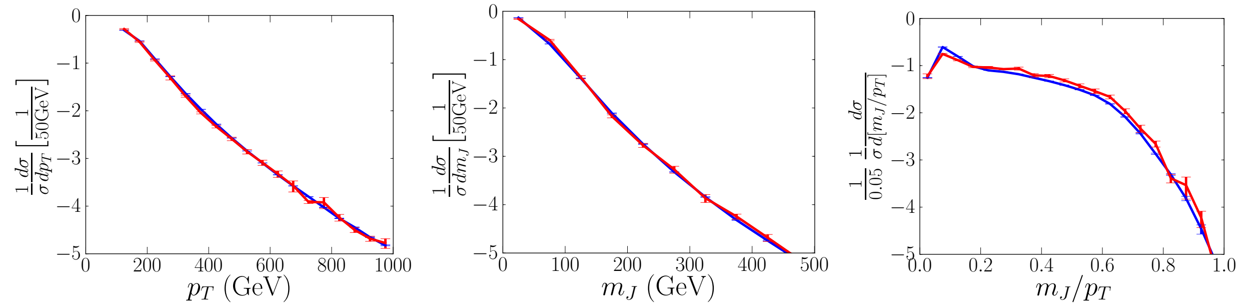

FIG. 4: Comparison between SHERPA (red) and MadGraph+PYTHIA (blue) for three relevant kinematic variables of the leading jet.

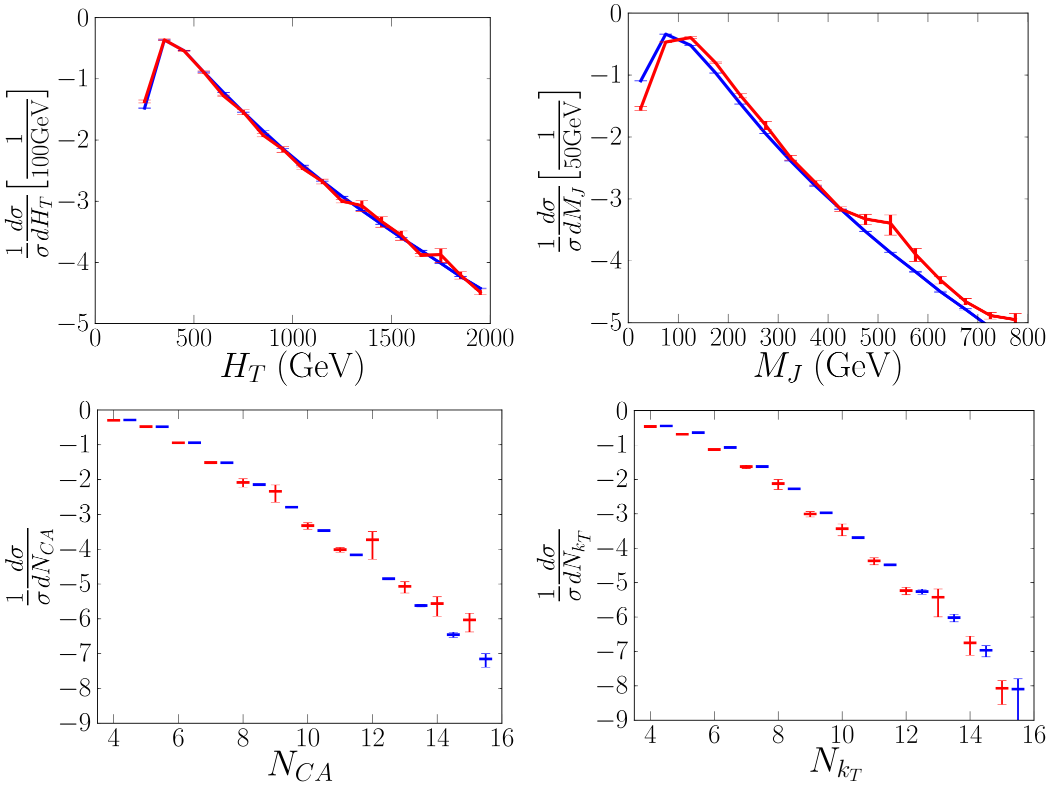

FIG. 5: Comparison between SHERPA (red) and MadGraph+PYTHIA (blue) for H_T (upper left), M_J (upper right), $N_{\rm CA}$ (lower left) and N_{k_T} (lower right). For the definition of the latter two observables see Sec. 2. For the bottom two distributions, the MadGraph+PYTHIA points have been shifted by half a unit to the right in order to facilitate the comparison.

3.3. Detector mockup and jet clustering

After showering, all hadron-level events are passed to the PGS 4 [37] detector simulation, which parameterizes the detector response. The detector parameters used are those of the default ATLAS PGS card. The PGS output is clustered into 0.1×0.1 cells in $\eta - \phi$ space, and then each cell is represented as a massless four-vector pseudoparticle. Finally, those pseudoparticles with rapidities $|y| < 2.5$ are fed into FastJet 3, which we use for jet clustering [16].

The primary purpose of PGS is to give estimates of the missing energy that arises from imperfect detectors. As we will see in Sec. 4.1, for the QCD backgrounds this is best accomplished by parametrizing the PGS QCD missing energy spectrum in terms of template functions. In Sec. 4.1.1, our Monte Carlo background calculations are compared against published ATLAS data. There we will see that the QCD missing energy templates will need to be rescaled to obtain a better fit to the data. PGS is also useful for simulating lepton identification efficiencies. However, for the primary proposal of this paper, no lepton requirements or vetoes are made, and so lepton identification efficiencies will not play a large role.

3.4. Treatment of leptons

The goal of this paper is to develop a search strategy that can dramatically reduce Standard Model backgrounds while making the least number of assumptions about the characteristics of the final state apart from its having a high multiplicity. Consequently, while it may be advantageous to require or veto on something like b-tagged jets or isolated leptons for a particular signal model, we do not do so here. For the broad class of signals we are interested in probing, the high final state multiplicity may be exclusively hadronic in origin or it may include some number of leptons. For models with multiple possible cascade topologies (or indeed *any* with top quarks or electroweak gauge bosons in the final state) both the hadronic and semi-hadronic modes may be simultaneously present, and signal discovery may require sensitivity to both channels. Consequently throughout this study we treat leptonic energy as hadronic energy. That is to say that in both fat jet clustering and subjet counting, hadronic and leptonic energy are treated democratically. This helps to ensure that signal efficiencies are not unnecessarily degraded. It is interesting to ask whether alternative treatments of the leptons might lead to effective search strategies without having to sacrifice the relative inclusiveness of the present search. Doing so, however, lies outside the scope of this paper.

4. DATA DRIVEN BACKGROUNDS

High multiplicity searches push into regions of phase space that are challenging to model, particularly in the case of the pure QCD backgrounds. For this reason it is important to have as many handles on the backgrounds as possible. In particular a data driven extrapolation of the background from a control region to the signal region would be especially valuable for corroborating background estimates available from Monte Carlo. In this section we explore how such a data driven estimate might be made. In Sec. 4.1 we specialize to the particular case of missing energy. The resulting missing energy templates allow us to achieve significantly improved statistics in our Monte Carlo estimates of QCD missing energy acceptances. In Sec. 4.1.1 we compare our Monte Carlo background estimates to published ATLAS data. Finally in Sec. 4.2 we discuss the possibility of extending the template method to take into account M_J and N_J. These latter results are preliminary and are not used for the background estimates that enter in the expected limits in Sec. 5.

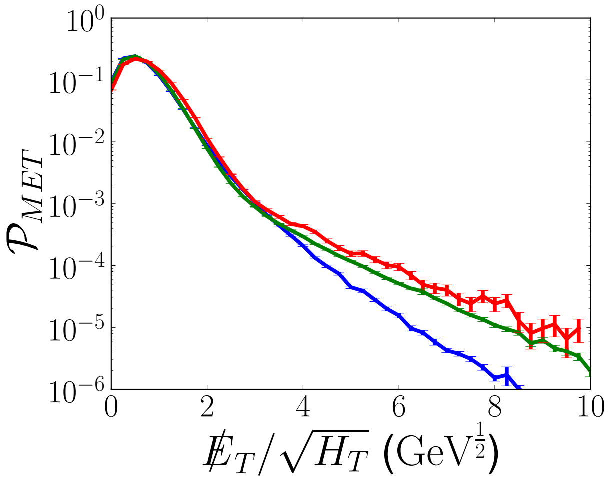

FIG. 6: The missing energy significance, $y = \not{E}_T/\sqrt{H_T}$ (GeV$^{\frac{1}{2}}$) in three different H_T bins: [300, 600] GeV (blue), [900, 1200] GeV (green), and [1500, 1800] GeV (red).

4.1. QCD missing energy templates

One of the purposes of this work is to reduce the dependence of new physics searches on missing energy requirements, which are particularly effective in reducing QCD backgrounds. Thus it is important that we model $\not{E}_T$, and in particular QCD $\not{E}_T$, as accurately as possible.

QCD missing energy typically arises from two distinct sources at the LHC. The first is from neutrinos lost in semi-leptonic decays of bottom and charm quarks. This irreducible form of missing energy gives a long non-Gaussian tail to missing energy distributions but can be estimated through Monte Carlo calculations. The second form of missing energy arises from detector effects that result in particles being lost or otherwise mismeasured. This form of missing energy is usually parameterized as a response function on the jet-by-jet level and is typically Gaussian for several standard deviations. The typical amount of missing energy scales as the square root of the jet energy, although there is a small linear term that takes over at large jet energies. For QCD events it is this latter form of missing energy, which arises from detector effects, that is dominant.

This article uses the approach of creating a probability distribution function for the missing energy of QCD events as a function of H_T (c.f. Sec. 4.2). This allows for a huge reduction in the number of Monte Carlo events necessary for accurate background estimates in the presence of missing energy requirements. Because the detector response is orthogonal to other jet properties that will be used (and is in any case being parametrized by PGS) this approach should faithfully reproduce the results of a much larger Monte Carlo calculation.

Specifically, a probability distribution function for missing energy significance,

$$y \equiv \not{E}_T/\sqrt{H_T},$$

as a function of H_T is constructed from the unweighted MadGraph QCD sample:

$$\mathcal{P}_{\rm MET}\,(y; H_T) \qquad \text{with} \qquad \int_0^\infty dy\, \mathcal{P}_{\rm MET}(y; H_T) = 1 \tag{11}$$

Thus for the rest of this study (where SHERPA will be used for all our QCD background estimates), although the H_T distribution will be modeled by SHERPA the mapping $H_T \to \not{E}_T$

Signal Region			Z^0/γ^*+jets		$W^\pm$+jets		VV+jets		$t\bar{t}$		QCD, $\alpha = 0.8$	
n_j	$m_{\rm eff}$	$\not\!\!E_T/m_{\rm eff}$	MG4	ATLAS	MG4	ATLAS	MG4	ATLAS	MG4	ATLAS	SHERPA	ATLAS
2	1000	0.40	66.51	61.0 ± 5.3	41.38	31.0 ± 24.1	4.84	9.14 ± 3.97	18.59	11.6 ± 2.8	0.14	0.10 ± 0.10
2	1900	0.30	1.99	1.17 ± 0.38	1.35	0.52 ± 0.68	0.19	0.57 ± 0.52	0.53	0.13 ± 0.13	0.00	—
3	1300	0.30	9.25	10.2 ± 1.0	5.31	5.5 ± 5.5	0.99	1.9 ± 0.9	3.23	2.4 ± 0.9	0.01	0.03 ± 0.03
4	1000	0.30	9.93	11.6 ± 1.0	7.49	6.9 ± 6.9	1.25	1.0 ± 0.7	9.22	6.7 ± 1.2	0.05	—
5	1700	0.15	0.37	0.43 ± 0.19	0.42	—	0.06	0.14 ± 0.07	0.84	0.44 ± 0.28	0.18	0.07 ± 0.09
6	1300	0.25	0.18	0.10 ± 0.09	0.08	0.16 ± 0.19	0.03	—	0.54	0.34 ± 0.24	0.00	—
6	1000	0.30	0.18	0.34 ± 0.17	0.13	0.14 ± 0.22	0.05	—	0.64	0.43 ± 0.16	0.00	—

TABLE 2: Comparison between backgrounds and ATLAS data driven estimates [38]. The QCD backgrounds are generated with SHERPA as described in Sec. 3.2.2, whereas the others are generated with MadGraph + Pythia as described in Sec. 3.1.

will be modeled by a combination of PGS and MadGraph. We use MadGraph for this purpose because doing so yields much better statistics (the ability to target specific H_T bins is one advantage). This PDF can be used event-by-event to determine the probability that an event, e_i, passes a specific $\not\!\!E_T$ requirement:

$$P(e_i) = \int_{\not\,E_{T\,\rm cut}}^{\infty} dy\, \mathcal{P}_{\rm MET}(y; H_{T\,e_i}) \tag{12}$$

Fig. 6 shows the missing energy significance distributions for various H_T windows. Note that these PDFs are just another way of parametrizing PGS's detector response. We will see below, in Sec. 4.1.1, that it will be necessary to scale these PDFs to ensure a better fit to published ATLAS data.

Looking forward to Sec. 4.2, the factorization we have introduced is given succinctly by

$$d\sigma(\not\!\!E_T, H_T) = \mathcal{P}_{\rm MET}(\not\!\!E_T/\sqrt{H_T}; H_T)\, d\sigma(H_T) \tag{13}$$

which can be compared to the analogous expression in Eq. 18 below.

4.1.1. Background validation

We validated our backgrounds against the signal regions used by an ATLAS search for squarks and gluinos [38]. This lead us to rescale our $\not\!\!E_T$ PDFs (see Sec. 4.1 and below), since the ATLAS PGS parameters were resulting in too large QCD $\not\!\!E_T$ acceptances. The results after the rescaling are shown in Table 2. Overall, our Monte Carlo background estimates agree with the data driven estimates given in the ATLAS study. In those cases where there is disagreement, our Monte Carlo always overestimates the ATLAS estimates so that the expected sensitivities presented in Sec. 5 should be reasonably conservative.

Each of the ATLAS signal regions includes a lepton veto. Because ATLAS lepton identification is only approximately reproduced by PGS, we expect larger differences between Monte Carlo estimates and ATLAS estimates for those backgrounds with leptons. This is indeed what we find in Table 2, where backgrounds with no leptons (like Z^0/γ^*+jets) match the data well. For events with $\not\!\!E_T$ arising from the decay of $W^\pm$ bosons, such as $W^\pm$+jets and $t\bar{t}$+jets, the comparisons are systematically off because PGS is not identifying isolated leptons with sufficiently high efficiency. This is not a major issue in the validation because the searches described in this article neither require nor veto on leptons.

As mentioned above, the PGS treatment of angular and energy smearing does not faithfully reproduce the response of the ATLAS detector. In particular, in the presence of $\not{E}_T$ cuts the ATLAS PGS parameters result in an overestimate of QCD backgrounds by an order of magnitude solely due to the treatment of $\not{E}_T$. In order to better reproduce the QCD backgrounds, the $\not{E}_T$ templates are rescaled by making the replacement

$$\mathcal{P}_{\text{MET}}\left(\frac{\not{E}_T}{\sqrt{H_T}}; H_T\right) \to \mathcal{P}_{\text{MET}}\left(\frac{1}{\alpha}\frac{\not{E}_T}{\sqrt{H_T}}; H_T\right) \tag{14}$$

The best fit to ATLAS estimates for QCD backgrounds in $\not{E}_T$-rich regions is with $\alpha = 0.8$. This rescaled $\not{E}_T$ template leads to significantly improved agreement between the Monte Carlo and ATLAS estimates of the QCD background.

4.2. Fat jet templates

The main obstacle to modeling 4-jet QCD production is the large dimensionality of the space of observables under consideration. The quantity we would like to understand is the 9-dimensional 4-(fat)jet differential cross section $d\sigma_{4\text{J}}(\not{E}_T, m_i, n_i)$. Here $\not{E}_T$ is the missing energy of the 4-jet event, the m_i are the masses of the four jets, and the n_i are the four subjet counts (using e.g. the n_{k_T} algorithm defined in Sec. 2). With $d\sigma_{4\text{J}}$ in hand the chosen cuts on $\not{E}_T$,

$$M_J \equiv \sum_i m_i \qquad \text{and} \qquad N_J \equiv \sum_i n_i \tag{15}$$

can be imposed, thus yielding the expected QCD background in the signal region.

To make progress it is useful to reduce the dimensionality of the problem. This can be done by making the *assumption* that each of the four jets is governed by a universal probability distribution

$$\rho_J(x, n; p_T) \tag{16}$$

with

$$x \equiv m/p_T$$

which describes the probability of a fat jet having n subjets and a particular value of m/p_T, as a function of the fat jet p_T. The mass of a fat jet is correlated with the number of subjets it contains, since it is impossible to get multiple subjets without having a sizable mass; consequently ρ_J does not factorize and it is necessary to construct a two dimensional PDF.

The assumption of universality is not completely valid, for one reason because quark-initiated jets and gluon-initiated jets will have different distributions. The hope is that ensembles of jets will have similar ratios of quark- versus gluon-initiated jets and that the distribution functions will not be radically different. In practice, this is the case since it is challenging to distinguish quark-initiated jets from gluon-initiated jets and since it is difficult to construct selection criteria that isolate one from the other. The assumption of universality is even more aggressive, however, since it implies that these distribution functions are independent of their environment. This assumption is known to be violated to some degree, particularly as jets come closer together, but the pull of the environment on properties of fat jets tends to be less than $\mathcal{O}(10\%)$ in magnitude.[5] In this section we will be satisfied to check these assumptions empirically with Monte Carlo calculations, leaving a more detailed study to future work. Note that we will not be applying the resulting background estimates when calculating the estimated sensitivity of our search strategy: the results presented in Sec. 5 will use the `SHERPA` Monte Carlo calculations described in Sec. 3.2.2. The results in this section are a first attempt at studying more aggressive uses of data driven approaches to QCD backgrounds.

The assumption underlying the form of this jet template is that a jet's substructure (e.g. its mass) is determined by its p_T and is independent of other jets in the event. The full 4-jet distribution is then obtained via the product:

$$d\sigma_{4\mathrm{J}}(\not{E}_T, m_i, n_i; p_{Ti}) = d\sigma_{4\mathrm{J}}(\not{E}_T, p_{Ti}) \prod_{i=1}^{4} \rho_J(x_i, n_i; p_{T\,i}) \tag{17}$$

Here the p_{Ti} are the transverse momenta of the four jets. Thus the 9-dimensional distribution $d\sigma_{4\mathrm{J}}(\not{E}_T, m_i, n_i)$ has been re-expressed as a function of the 5-dimensional distribution $d\sigma_{4\mathrm{J}}(\not{E}_T, p_{Ti})$ and the 3-dimensional jet template $\rho_J(x, n; p_T)$. A further reduction can be made by assuming that $\not{E}_T$ only depends on the quantity $H_T \equiv \sum p_{Ti}$. With this assumption we can introduce the $\not{E}_T$ template $\mathcal{P}_{\mathrm{MET}}(y; H_T)$, with

$$y \equiv \not{E}_T / H_T^{\frac{1}{2}}$$

thus ending up with the factorization

$$d\sigma_{4\mathrm{J}}(\not{E}_T, m_i, n_i; H_T, p_{Ti}) = d\sigma_{4\mathrm{J}}(p_{Ti}) \mathcal{P}_{\mathrm{MET}}(y; H_T) \prod_{i=1}^{4} \rho_J(x_i, n_i; p_{T\,i}) \tag{18}$$

Ultimately it is an experimental question whether such a factorization holds. At some level we certainly expect correlations between the four jets and deviations from the form of Eq. 18. For example, we would expect correlations to arise from color (re)connections as well as out-of-jet radiation. The presence of significant pile-up (so long as it remains unsubtracted) would also tend to result in (positive) correlations between the jets.

In the case that the correlations are large it may be necessary to systematically include corrections to Eq. 18. We anticipate that some kind of principal component analysis or form of tensor decomposition would be applicable. We leave this interesting question to future work. For the remainder of this section we would like to explore the degree to which the universality assumptions underlying Eq. 18 are valid in the only data sample available to us, namely the 4.8 million `SHERPA` events described in Sec. 3.2.2.

In a realistic experimental study one would presumably want to measure $d\sigma_{4\mathrm{J}}(p_{Ti})$ and $\rho_J(x, n; p_T)$ from independent samples. Given our somewhat limited statistics, however, we will instead 'measure' $d\sigma_{4\mathrm{J}}(p_{Ti})$ and $\rho_J(x, n; p_T)$ from the same 4-jet sample and use Eq. 18 to construct an estimate of the full 9-dimensional distribution. This will allow us to estimate acceptances after imposing $\not{E}_T$, M_J and N_J cuts. The degree to which this

[5] See e.g. Figure 2 in ref. [13].

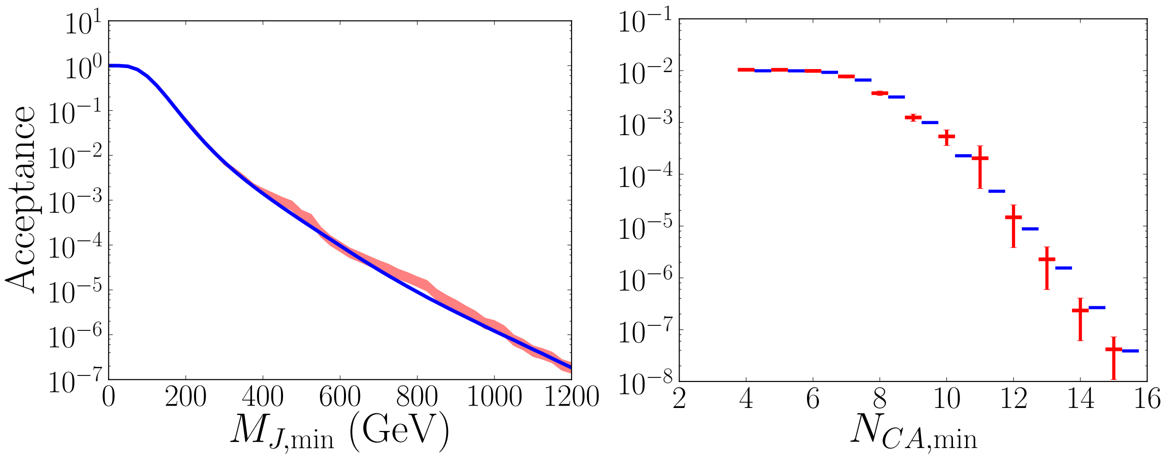

FIG. 7: Testing the jet template ansatz. The figure on the left compares the raw M_J cut acceptance (red) to the template estimate (blue). The figure on the right is analogous, with a sliding $N_{\rm CA}$ cut and a fixed cut $M_J > 280$ GeV. The red uncertainties are statistical (the statistical uncertainties for the template estimate are not shown, since by construction they are parametrically smaller than the raw uncertainties).

procedure reproduces cut acceptances in the raw event sample will reflect the viability of the jet template and $\not{E}_T$ template ansätze.

Given our somewhat limited statistics, it is difficult to judge whether deviations between the raw and template cut acceptances (see Fig. 7) are an indication of deviations from Eq. 18 or just statistical fluctuations. Nevertheless, the approximate agreement over 7 orders of magnitude of cut acceptance in Fig. 7 is a promising result, although more sophisticated statistical methods would likely be required for a robust experimental analysis.

When the jet template and $\not{E}_T$ template ansätze are appropriate, they have the advantage of reducing the statistical uncertainties in $d\sigma_{\rm 4J}(\not{E}_T, m_i, n_i)$. This follows directly from the reduced dimensionality of the problem. This reduction is especially significant in the tails of the distribution, where statistical uncertainties are parametrically reduced by virtue of the fact that—due to the convolution—they receive sizeable contributions from statistically rich parts of the component probability distributions. For example, a rare 4-jet event with $N_J = 12$ will be dominated by contributions from four jets with three subjets each. The probability of a single fat jet having three subjets at a given m/p_T and p_T can be measured (or calculated) readily.

Note that in the above we have discussed template methods in the context of LHC data. Another possibility is to use template methods to extend Monte Carlo calculations—indeed this is precisely what we did in the previous section for the specific case of missing energy. In the context of Monte Carlo template methods have the obvious advantage of reducing statistical uncertainties in the tails of distributions. They also offer the possibility of extending lower multiplicity calculations to a higher multiplicity regime in the following sense. Take for example `SHERPA`, which can generate up to $n_j = 6$ jets in the final state at the matrix element level. Thus a `SHERPA` dijet sample will include jets with up to five partons generated through matrix elements. When a fat jet template obtained from such a dijet sample is combined with a 4-jet distribution $d\sigma_{\rm 4J}(p_{Ti})$, the resulting distributions will extend the nominal reach of `SHERPA` beyond $n_j = 6$. Of course, the theoretical validity of such a method is a delicate matter. Given the importance of having reliable Monte Carlo estimates in the tails of distributions, however, such an approach deserves further study. For example, it would be important to investigate correlations between fat jets. Since SHERPA extends to $n_j = 6$, one would be able to, among other possibilities, study 3-jet events and observe what happens as fat jets come closer together.

5. RESULTS

This section investigates the benefit of incorporating a subjet counting observable, namely N_J, into high multiplicity searches based off the summed jet mass observable, M_J. Sec. 5.1 discusses the models used to quantify the improvement in searches that results from incorporating N_J. These are supersymmetric models whose phenomenology involves the pair production of gluinos that subsequently decay into the lightest supersymmetric particle. Both R-parity conserving and R-parity violating models are considered. We choose these benchmark signals because they are well known in the literature and are easy to implement in Monte Carlo event generators. Sec. 5.2 describes the signal and background distributions in signal-like regions. Sec. 5.3 describes the criterion that was used to create optimized search regions for the benchmark signals. Sec. 5.4 describes the expected sensitivity of the optimized search regions to the benchmark signals. Finally, Sec. 5.5 compares the optimized search regions to previous searches.

5.1. Benchmark signals

The goal of this work is to gain access to a large class of signals without specifically targeting any one signal. Nevertheless, it is useful to have some benchmark models to consider. While these benchmark models are plausible extensions of the Standard Model, more than anything else they are meant to exhibit features of theories that produce high multiplicity final states. For any single theory, there are numerous handles beyond large multiplicity that could allow for additional discrimination between signal and background. For instance, many of our benchmark signals have leptons and b-jets. These are powerful handles that can be used in conjunction with the methods in this article, but they are not generic to all signals that produce high multiplicity final states. Therefore, these additional handles will not be used.

The eight benchmark models considered in this article arise from the pair production of gluinos. These benchmarks provide relatively straightforward ways of toggling the multiplicity of final state partons within a class of models that is easily implemented in all the standard Monte Carlo calculation packages. Each benchmark model is generated using MadGraph 4.5.1 [25–27] and with up to two additional jets:

$$pp \to \tilde{g}\tilde{g} + n_{\tilde{g}} j \qquad \text{with} \qquad \mathrm{n}_{\tilde{\mathrm{g}}} \leq 2 \tag{19}$$

The MLM matching scheme with a shower-$k_\perp$ scheme was used to account for the extra radiation. The events were showered and hadronized using Pythia 6.4 [28]. K factors for the signals were calculated using Prospino 2.1 [39].

A collection of signals with diverse phenomenology is considered in order to better explore/delineate the efficacy of the subjet techniques used in this paper. This diversity arises through the variety of gluino decay topologies that are possible. The gluino can decay to light or heavy quarks; it can decay directly to the LSP or instead through a cascade. The

Model	Gluino Decay		Electroweakino		LSP Decay		Final State Partons
	$q\bar{q}\chi(+4)$	$t\bar{t}\chi_i(+12)$	χ_0	$\chi_2(+8)$	Stable$(+0)$	$cbs(+6)$	
$\mathcal{G}_0$	✓		✓		✓		4
$\mathcal{G}_1$		✓	✓		✓		12
$\mathcal{G}_2$	✓			✓	✓		12
$\mathcal{G}_3$		✓		✓	✓		20
$\mathcal{G}_4$	✓		✓			✓	10
$\mathcal{G}_5$		✓	✓			✓	18
$\mathcal{G}_6$	✓			✓		✓	18
$\mathcal{G}_7$		✓		✓		✓	26

TABLE 3: The eight benchmark signals used in this paper. The numbers in parentheses indicate the number of final state partons added by choosing that particular branch of the decay topology.

theory can be R-parity conserving or R-parity violating. Decay topologies with cascade decays or decays involving top quarks as well as certain RPV topologies will lead to very high multiplicity events with 12 or more final state partons. Indeed one benchmark signal we consider ($\mathcal{G}_7$) has a spectacular 26 final state partons. The expected sensitivities to all the signals presented here will be given in Sec. 5.4.

In more detail, all the processes outlined here start with a gluino decaying to either a light quark or a top quark pair and a neutralino:

$$\tilde{g} \to q\bar{q}\chi \qquad \text{or} \qquad \tilde{g} \to t\bar{t}\chi \tag{20}$$

The neutralino χ may be the LSP χ_0 or one of the heavier electroweakinos. For simplicity, only decays to the LSP and to the NNLSP χ_2 are considered, where the latter decay chain results in a 2-step cascade:

$$\chi_2 \to V\chi_1 \to VV'\chi_0 \tag{21}$$

Finally the LSP may or may not decay into jets. The constraints on R-parity violation are much weaker for decays into heavy flavor. If the LSP is lighter than 200 GeV, then the decays will be dominantly with the $\lambda_{ijk}U_i^c D_j^c D_k^c$ flavor structure $(ijk) = (2, 3, 2)$. The resulting decay topology is

$$\chi_0 \to qqq = c\,b\,s. \tag{22}$$

If the mass of the LSP χ_0 is above 200 GeV, then it is possible for the dominant R-parity violating decay mode to be tbs, resulting in four more final state partons over the cbs decay mode. To keep the number of benchmark signals to a manageable number, we do not include this decay mode in any of our benchmarks. All of these possibilities taken together result in eight different gluino decay topologies that span a range of final state parton multiplicities, see Table 3.[6]

[6] Here we use the term parton in a loose sense that includes the leptons from gauge boson decays.

For all signals, we choose the LSP mass using the formula

$$m_{\chi_0} = m_{\tilde{g}}/10 \tag{23}$$

For the signal models involving heavier electroweakinos we choose the intermediate masses as follows:

$$m_{\chi_2} = (m_{\tilde{g}} + m_{\chi_0})/2 \qquad m_{\chi_1} = (m_{\chi_2} + m_{\chi_0})/2 \tag{24}$$

Note that if the gluinos decay to light quarks and the LSP, the final state will have between 4 (RP conserving) and 10 (RPV) partons, which will make these topologies hard to discriminate against the $t\bar{t}$ background, especially in the former case. In the case of cascade decays or decays involving top quarks, however, there will be at least 12 partons in the final state and the method outlined in this article should prove more effective. Moreover, if R-parity is violated, each LSP will decay to three quarks, thus adding 6 jets to the final state. Cuts on the total number of subjets could then provide a competitive replacement for MET cuts for these kinds of signals.

These signals are simply meant as benchmark models to test the sensitivity of our search to high multiplicity final states. The search presented here should prove effective for any signal implying the existence of final states with 8 or more final state jets.

5.2. Distributions for signals and backgrounds

One of the goals of this paper is to investigate the degree to which $\not{E}_T$ cuts for the signals of interest can be substituted (more realistically, loosened) by requiring particular jet substructure. That this is challenging can be inferred from Fig. 8, which shows the $\not{E}_T$ distributions of the various backgrounds with the $\not{E}_T$ distributions of two benchmark signals superimposed. Although the dominant QCD background is significantly reduced by a $\not{E}_T$ cut of order 150 GeV, any loosening of this cut dramatically increases the number of QCD events in the search region.

The remaining backgrounds, most of which have intrinsic $\not{E}_T$, require additional cuts to be suppressed. As shown in Fig. 8, a cut on the sum of the jet masses of order 300 GeV is effective. That a cut on M_J does not exhaust the discrimination available from jet substructure can be seen in Fig. 9, which illustrates how even at large values of M_J the N_{CA} and N_{kT} distributions of a typical high multiplicity signal are well separated from that of the QCD background. The observation of similar behavior for the N-subjettiness ratio τ_3/τ_2 [40] suggests that this separation should hold for real QCD data as well.

Thus N_J cuts should be complementary to M_J cuts, allowing for the possibility that $\not{E}_T$ cuts could be loosened. This complementarity is made more explicit in the bottom row of Fig. 8, which shows the distributions of N_{CA} and N_{kT} for the various backgrounds and two benchmark signals after imposing $\not{E}_T$ and M_J cuts. Once these cuts have been imposed, the dominant remaining background comes from $t\bar{t}$+jets (and to a lesser extent V+jets), as it has the largest number of final state partons. In order for $t\bar{t}$+jets to pass the $\not{E}_T$ cut, the $W^{\pm}$ bosons have to decay semi-leptonically; while to pass the M_J cut, the final state partons must be distributed in phase space such that they form massive jets upon fat jet clustering. The combination of all these cuts strongly suppresses the various backgrounds.

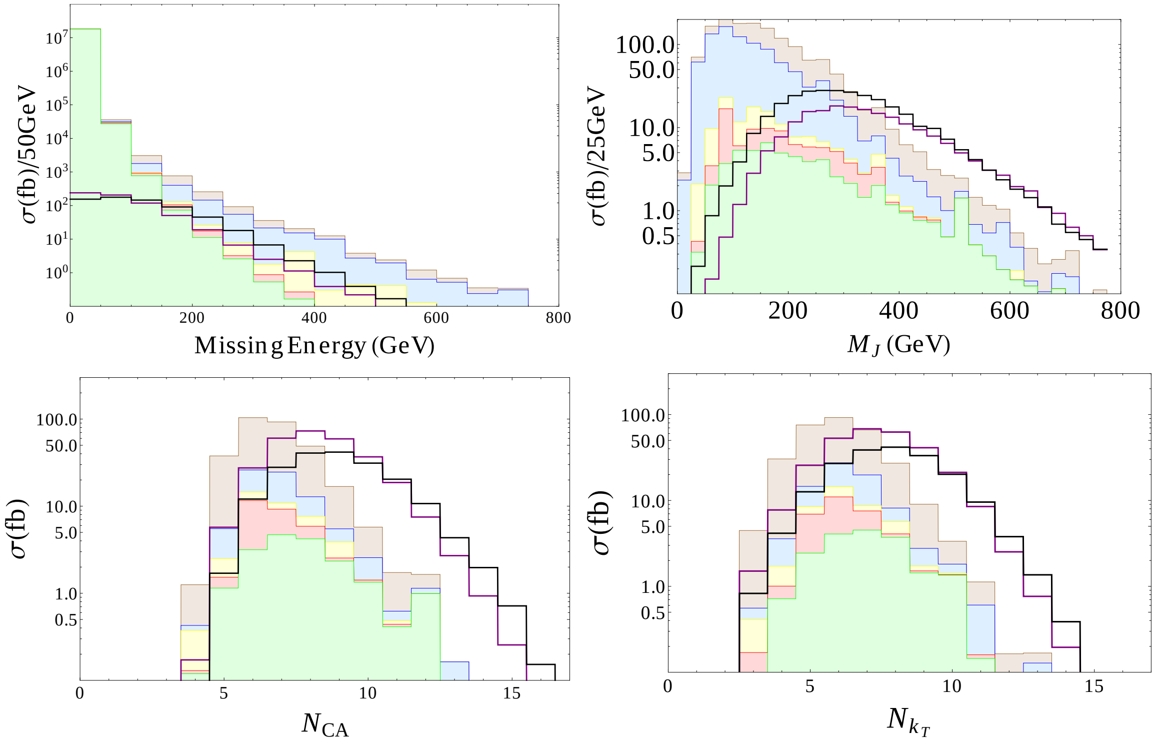

FIG. 8: *Top Left*: $\not{E}_T$ distributions for signals and backgrounds after requiring four or more fat jets. *Top Right*: M_J distributions, after requiring four or more fat jets and $\not{E}_T > 150$ GeV. *Bottom Left*: $N_{\rm CA}$ distribution after requiring four or more fat jets, $M_J > 280$ GeV and $\not{E}_T > 150$ GeV. *Bottom Right*: $N_{\rm kT}$ distribution after requiring four or more fat jets, $M_J > 280$ GeV and $\not{E}_T > 150$ GeV. Stacked histograms show the SM backgrounds, which include top and single top (light brown), $V + nj$ (light blue), diboson (light yellow), QCD (light green), and the remaining non-QCD backgrounds mentioned in Sec 3.1 (light red). The distributions for a 600 GeV gluino in the $\mathcal{G}_1$ and $\mathcal{G}_3$ topologies are shown in purple and black, respectively. Note that the $N_{\rm CA}$ and $N_{\rm kT}$ distributions for $\mathcal{G}_3$ (with 20 final state partons) are not substantially different from $\mathcal{G}_1$ (with 12).

5.3. Optimizing search strategies

The simplified models introduced in Sec. 5.1 can be used to develop broad search strategies that cover the model space. This section describes the method that was used to construct the minimal number of signal regions necessary to cover the entire space of simplified models. The method used was introduced in ref. [41], developed further further in ref. [42], and is based off the variable "efficacy," which is defined below.

In order to demonstrate the usefulness of N_J cuts, we present two separately optimized search strategies. The first uses only M_J and $\not{E}_T$ cuts, while the second uses N_J, M_J and $\not{E}_T$ cuts. Since the first set of searches is a subset of the second, the second will always do better. The degree to which the more complex search strategy can be judged superior (if at all) will depend on the resulting sensitivities, the number of search regions required, and the sorts of cuts favored by the introduction of N_J.

The two search strategies are defined as

$$\hat{\mathcal{C}} = \{(\not{E}_{T\,\rm min}, M_{J\,\rm min})\} \qquad \text{and} \qquad \mathcal{C} = \{(\mathrm{N_{subjet\,min}}, \not{E}_{\rm T\,min}, \mathrm{M_{J\,min}})\}. \tag{25}$$

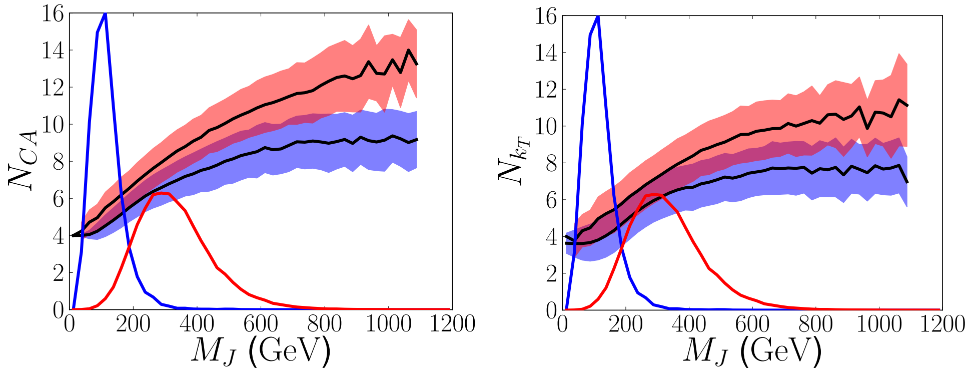

FIG. 9: $N_{\rm CA}$ (left) and N_{k_T} (right) versus M_J for signal topology $\mathcal{G}_4$ with $m_{\tilde{g}} = 600\,$GeV (red band) and QCD background (blue band). Each band illustrates $\pm 1\sigma$ standard deviation about the mean, which is denoted by a black line. The superimposed M_J distributions are for the signal (red) and the QCD background (blue). As defined in Sec 2, N_{CA} and N_{k_T} are the sum of the number of subjets of the four leading jets of each event. The events considered are required to have at least four jets with $p_T > 50\,$GeV and the p_T of their leading jet must be greater than 100 GeV.

where the values of each of the cut requirements are taken from the following sets:

$$\begin{aligned} N_{J\,\rm min} &\in \{0, \ldots, 16\} \\ \not{E}_{T\,\rm min} &\in \{0, 25, \ldots, 600\}\,\text{GeV} \\ M_{J\,\rm min} &\in \{0, 25, \ldots, 1600\}\,\text{GeV}\,. \end{aligned}$$

This results in 1625 search regions for $\hat{\mathcal{C}}$ and 27,625 search regions for $\mathcal{C}$. The optimized search strategies will make use of only a small subset of these search regions.

A given signal region or set of cuts, C_i, will yield an expected limit on the cross section times branching ratio, $\sigma \times \mathcal{B}$, for a given simplified model at the 95% C.L. given by

$$(\sigma \times \mathcal{B})_i \;=\; \frac{\Delta(B)_i}{\mathcal{L} \times \epsilon(M)_i} \tag{26}$$

Here $\epsilon(M)_i$ is the efficiency of C_i for the model M and $\mathcal{L}$ is the integrated luminosity, while $\Delta(B)_i$ is the maximum number of allowed signal events at the 95% C.L. if B background events are expected after the cuts and in fact fit the data. We take

$$\Delta(B) \;=\; 2 \times \sqrt{\text{Stat}(B)^2 + (\epsilon_{\rm syst} B)^2}, \tag{27}$$

where Stat(B) is the Poisson limit on B and $\epsilon_{\rm syst}$ is the systematic uncertainty. Throughout we will take $\epsilon_{\rm syst} = 30\%$. The Monte Carlo statistical uncertainties in B and $\epsilon(M)_i$, i.e. δB and $\delta\epsilon(M)_i$, are taken into account by making the replacements

$$B \to B + \delta B \qquad \text{and} \qquad \epsilon(\mathrm{M})_\mathrm{i} \to \epsilon(\mathrm{M})_\mathrm{i} - \delta\epsilon(\mathrm{M})_\mathrm{i} \tag{28}$$

which result in conservative limits.

The optimal limit on a model M is then given by

$$(\sigma \times \mathcal{B})_{\rm opt} \;=\; \{\min((\sigma \times \mathcal{B})_i) : i \in \{1, N_{\rm cuts}\}\}\,, \tag{29}$$

where the number of search regions is $N_{\rm cuts} = 1625$ or $27,625$ depending on whether N_J cuts are being used. It is natural to quantify the "goodness" of a cut C_i by how close it is to optimal. For this purpose, we introduce the efficacy of a cut

$$\mathcal{E}(C_i) \;=\; \frac{(\sigma\times\mathcal{B})_i}{(\sigma\times\mathcal{B})_{\rm opt}}. \tag{30}$$

This is the ratio of the expected limit on the production cross section using a particular cut C_i divided by the expected limit on the cross section using the optimal set of cuts. An efficacy of 1.0 is ideal. Thus the best search strategy for covering a collection of model points $\{M_j\}$ will be a combination of cuts $\{C_i\}$ such that $\mathcal{E}$ is close to one for each M_j for at least one of the C_i. This article will use $\mathcal{E} \leq \mathcal{E}_{\rm crit} = 1.5$ as the criterion for optimizing the number of search regions. That is, each model point M will be covered by at least one cut that yields a limit on $\sigma\times\mathcal{B}$ that is within a factor of 1.5 of the optimal limit. The efficacy approach has several advantages:

- it ensures near optimal coverage over the range of signals;
- it allows for a fair comparison between different sets of observables;
- it allows for a reasonable comparison to the ATLAS high multiplicity search, which makes use of 6 search regions; and
- each signal is grouped with like signals on the basis of which search region it is covered by.

Finding a search strategy that covers all models with a desired efficacy is computationally challenging because the configuration space is enormous, with $2^{N_{\rm cuts}}$ possible search strategies. Since a brute force search is not feasible, we use a genetic algorithm to construct the minimal set of search regions needed to cover the entire space of models. This algorithm, which we find to be quite effective for the task at hand, is described in App. A and is based off a genetic algorithm described in detail in [42].

5.4. Expected sensitivity

Expected sensitivities to the various benchmark signals at $\sqrt{s} = 8$ TeV are depicted in Figures 10-12. These are presented as expected 95% exclusion limits on $\sigma\times\mathcal{B}$ (the production cross section times the branching ratio into that particular gluino decay topology) as a function of the gluino mass and for an integrated luminosity of 30 fb^{-1}. As expected, the performance of the M_J + $\not{\!\!E}_T$ +N_J search depends strongly on the final state multiplicity as well as the intrinsic $\not{\!\!E}_T$ of the signal.

The results of the subjet counting search are best revealed by comparing the optimal search regions for the M_J + $\not{\!\!E}_T$ search to those of the M_J + $\not{\!\!E}_T$ +N_J search. For the case of $N_{\rm CA}$ the former has 6 search regions, while the latter has 5 search regions, see Tables 4 and 5. Interestingly, the efficacy criterion groups the signals into roughly similar signal classes, with the difference that some of the M_J and $\not{\!\!E}_T$ cuts move around once $N_{\rm CA}$ cuts are introduced.

The first class of signals consists of $\mathcal{G}_0$ alone, has intrinsic $\not{\!\!E}_T$ from the stable (non-RPV) LSP and only 4 final state partons. Consequently the cuts (and expected limits, see Fig. 10) do not change substantially after the introduction of a (trivial) $N_{\rm CA}$ cut.

Search Region			Models Covered		Background (for $30\,\mathrm{fb}^{-1}$)				
	M_J	$\not\!\!E_T$	Class	$m_{\tilde{g}}$	QCD	$t\bar{t}$	V+jets	Other	Total
1	1000	0	$\mathcal{G}_4$	$m_{\tilde{g}} \lesssim 1.0\,\mathrm{TeV}$	495 ± 61.5	2.38 ± 0.69	6.93 ± 2.73	0.13 ± 0.10	505 ± 62
2	1350	0	$\mathcal{G}_4$	$m_{\tilde{g}} \gtrsim 1.0\,\mathrm{TeV}$	13.7 ± 1.5	$\lesssim 0.1$	0.54 ± 0.54	$\lesssim 0.1$	14.3 ± 1.6
3	400	400	$\mathcal{G}_0$ $\mathcal{G}_1$	$m_{\tilde{g}} \lesssim 1.2\,\mathrm{TeV}$ $0.8\,\mathrm{TeV} \gtrsim m_{\tilde{g}} \gtrsim 1.1\,\mathrm{TeV}$	0.38 ± 0.04	16.63 ± 1.81	14.30 ± 2.62	4.40 ± 1.52	35.71 ± 3.53
4	500	200	$\mathcal{G}_1$ $\mathcal{G}_{2,3}$	$m_{\tilde{g}} \lesssim 0.8\,\mathrm{TeV}$ $m_{\tilde{g}} \lesssim 0.9\,\mathrm{TeV}$	23.9 ± 4.9	54.6 ± 3.3	28.0 ± 5.6	6.26 ± 1.52	112.8 ± 8.2
5	625	425	$\mathcal{G}_0$ $\mathcal{G}_1$ $\mathcal{G}_{2,3}$	$m_{\tilde{g}} \gtrsim 1.2\,\mathrm{TeV}$ $m_{\tilde{g}} \gtrsim 1.1\,\mathrm{TeV}$ $m_{\tilde{g}} \gtrsim 1.3\,\mathrm{TeV}$	0.09 ± 0.02	0.59 ± 0.34	0.73 ± 0.73	0.47 ± 0.29	1.89 ± 0.86
6	725	175	$\mathcal{G}_{2,3}$ $\mathcal{G}_{5,6,7}$	$0.9\,\mathrm{TeV} \lesssim m_{\tilde{g}} \lesssim 1.3\,\mathrm{TeV}$ all	5.28 ± 0.72	5.34 ± 1.03	2.85 ± 1.08	0.41 ± 0.18	13.87 ± 1.67

TABLE 4: Search regions for the M_J + $\not\!\!E_T$ search with cuts in GeV and assuming 30% systematic uncertainties. For each search region C_i the column 'Models Covered' lists the benchmark models that are optimally covered by C_i. The search regions are chosen using the efficacy criterion $\mathcal{E} < 1.5$. The background uncertainties shown are statistical.

Search Region				Models Covered		Background (for $30\,\mathrm{fb}^{-1}$)				
	M_J	$\not\!\!E_T$	N_{CA}	Class	$m_{\tilde{g}}$	QCD	$t\bar{t}$	V+jets	Other	Total
1	450	450	0	$\mathcal{G}_0$	all	0.18 ± 0.26	8.31 ± 1.28	2.05 ± 1.08	0.64 ± 0.26	11.18 ± 1.70
2	1050	0	13	$\mathcal{G}_4$	all	21.60 ± 3.03	$\lesssim 0.1$	$\lesssim 0.1$	0.03 ± 0.01	21.63 ± 3.03
3	475	275	11	$\mathcal{G}_1$ $\mathcal{G}_2$ $\mathcal{G}_3$	all $m_{\tilde{g}} \gtrsim 0.8\,\mathrm{TeV}$ $m_{\tilde{g}} \gtrsim 0.9\,\mathrm{TeV}$	0.96 ± 0.46	4.16 ± 0.91	0.78 ± 0.59	0.03 ± 0.01	5.90 ± 1.18
4	525	125	12	$\mathcal{G}_2$ $\mathcal{G}_3$ $\mathcal{G}_{5,6}$	$m_{\tilde{g}} \lesssim 0.8\,\mathrm{TeV}$ $m_{\tilde{g}} \lesssim 0.9\,\mathrm{TeV}$ $m_{\tilde{g}} \gtrsim 0.9\,\mathrm{TeV}$	7.86 ± 1.92	7.72 ± 1.24	6.71 ± 4.58	0.33 ± 0.19	22.65 ± 5.11
5	425	125	14	$\mathcal{G}_{5,6}$ $\mathcal{G}_7$	$m_{\tilde{g}} \lesssim 0.9\,\mathrm{TeV}$ all	1.08 ± 0.32	1.19 ± 0.49	$\lesssim 0.1$	0.01 ± 0.01	2.26 ± 0.58

TABLE 5: Search regions for the M_J + $\not\!\!E_T$ + N_{CA} search with M_J and $\not\!\!E_T$ cuts in GeV and assuming 30% systematic uncertainties. For each search region C_i the column 'Models Covered' lists the benchmark models that are optimally covered by C_i. The search regions are chosen using the efficacy criterion $\mathcal{E} < 1.5$. The background uncertainties shown are statistical.

The second class of signals consists of $\mathcal{G}_4$ alone, which differs from $\mathcal{G}_0$ in that the LSP undergoes the RPV decay $\chi \to cbs$. Consequently there is no intrinsic $\not\!\!E_T$, and both search strategies cover $\mathcal{G}_4$ with search regions that have trivial $\not\!\!E_T$ cuts. Since, however, $\mathcal{G}_4$ is a high multiplicity signal with 10 final state partons, the corresponding N_{CA} search region imposes a significant cut $N_{\mathrm{CA}} \geq 13$ with a loosened M_J cut. This results in an expected limit on $\sigma \times \mathcal{B}$ that is better by a factor 2 to 4 compared to the optimized M_J + $\not\!\!E_T$ search

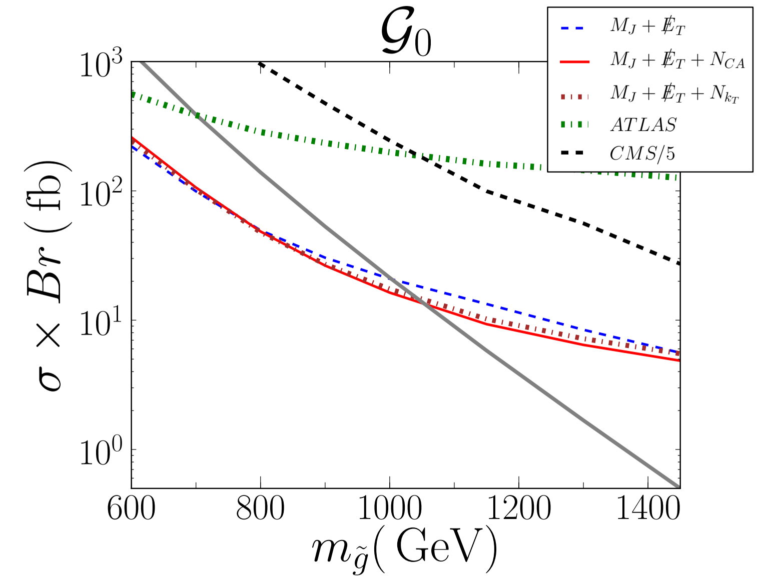

FIG. 10: 95% exclusion limits on $\sigma \times \mathcal{B}$ for the $M_J + \not{E}_T$ search (dashed blue), the $M_J{+}\not{E}_T{+}N_{\rm CA}$ search (solid red), and the $M_J{+}\not{E}_T{+}N_{\rm k_T}$ search (dash-dotted brown) for signal $\mathcal{G}_0$, which has only 4 final state partons. The exclusion limits given by the ATLAS high multiplicity search [43] (dash-dotted green line) and the CMS black hole search [44] (dashed black) as well as the NLO gluino production cross section (grey solid) are also shown. The systematic uncertainty on the background is assumed to be 30%. Note that the CMS limit is rescaled by a factor of 5.

(see Fig. 11) but that is nevertheless weaker than what would be needed to exclude the benchmark gluino cross section.

The third class of signals consists of $\mathcal{G}_5$, $\mathcal{G}_6$ and $\mathcal{G}_7$. These signals have intrinsic $\not{E}_T$ from top quarks or electroweak gauge bosons produced in the gluino decay chain. They also have especially large final state multiplicities, since the LSPs at the end of the decay chain end in the RPV decay $\chi \to cbs$. The inclusion of a cut $N_{\rm CA} \geq 12 - 14$ improves the expected limit by a factor of $2 - 5$ depending on the specific signal and gluino mass. The $\not{E}_T$ cut is loosened by $50\,\mathrm{GeV}$, while the M_J cut is lowered by $200 - 300\,\mathrm{GeV}$. This represents a modest success in trading our reliance on $\not{E}_T$ cuts for a more refined use of jet substructure observables.

The fourth and final class of signals consists of $\mathcal{G}_1$, $\mathcal{G}_2$ and $\mathcal{G}_3$. These signals have large intrinsic $\not{E}_T$ because the LSPs at the end of the gluino decay chain are stable and because top quarks and/or electroweak gauge bosons are produced in the decay chain. The top quarks and electroweak gauge bosons also ensure that the final state multiplicity is high. The inclusion of a $N_{\rm CA}$ cut improves the expected limit on $\sigma \times \mathcal{B}$ by a factor $2 - 4$ for low and intermediate gluino masses, with little or no improvement at large gluino masses. The inclusion of a $N_{\rm CA}$ cut also loosens (in most places) the requirements on M_J and $\not{E}_T$ by $50 - 200\,\mathrm{GeV}$ and $50 - 150\,\mathrm{GeV}$, respectively. This demonstrates that for these signals with significant $\not{E}_T$ there is more room for loosening $\not{E}_T$ requirements in favor of N_J requirements.

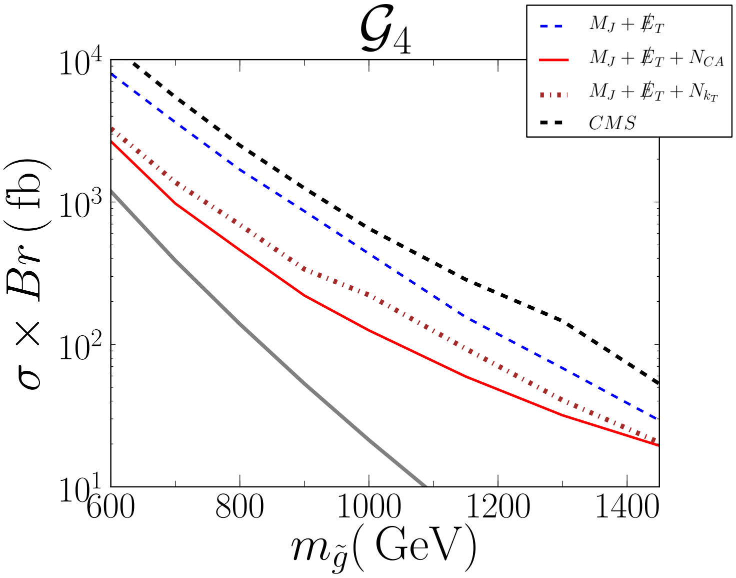

FIG. 11: 95% exclusion limits on $\sigma \times \mathcal{B}$ for the $M_J + \not{E}_T$ search (dashed blue), the $M_J + \not{E}_T + N_{\text{CA}}$ search (solid red), and for the $M_J + \not{E}_T + N_{\text{k}_\text{T}}$ search (dash-dotted brown) for signal $\mathcal{G}_4$, which has 10 final state partons and no intrinsic $\not{E}_T$. The exclusion limit given by the CMS black hole search [44] (dashed black) as well as the NLO gluino production cross section (grey solid line) are also shown. The systematic uncertainty on the background is assumed to be 30%. The ATLAS limit is not shown because it is orders of magnitude worse than the others due to its strict $\not{E}_T$ requirement.

5.5. Comparison with previous searches

This section presents a comparison of the techniques proposed in this article to previous searches. This is not meant to be a complete survey of all the searches that have been performed at the LHC and that are sensitive to high multiplicity signals. Two searches are considered. The first, presented in Sec. 5.5.1, is an ATLAS search that requires up to 9 $R = 0.4$ jets with missing energy. The second search, presented in Sec. 5.5.2, is a search for "black holes" at CMS. These two searches are different attempts at gaining access to high multiplicity final states. We find that the methods presented in this article are competitive.

5.5.1. ATLAS High Multiplicity Search

A comparison is made with ATLAS's most up-to-date high multiplicity search, which makes use of $5.8\,\text{fb}^{-1}$ at 8 TeV [43]. This search clusters events into $R = 0.4$ anti-k_T jets and looks at 6 search regions:

- $n_j \geq 7$ with $p_T \geq 55\,\text{GeV}$
- $n_j \geq 8$ with $p_T \geq 55\,\text{GeV}$
- $n_j \geq 9$ with $p_T \geq 55\,\text{GeV}$
- $n_j \geq 6$ with $p_T \geq 80\,\text{GeV}$

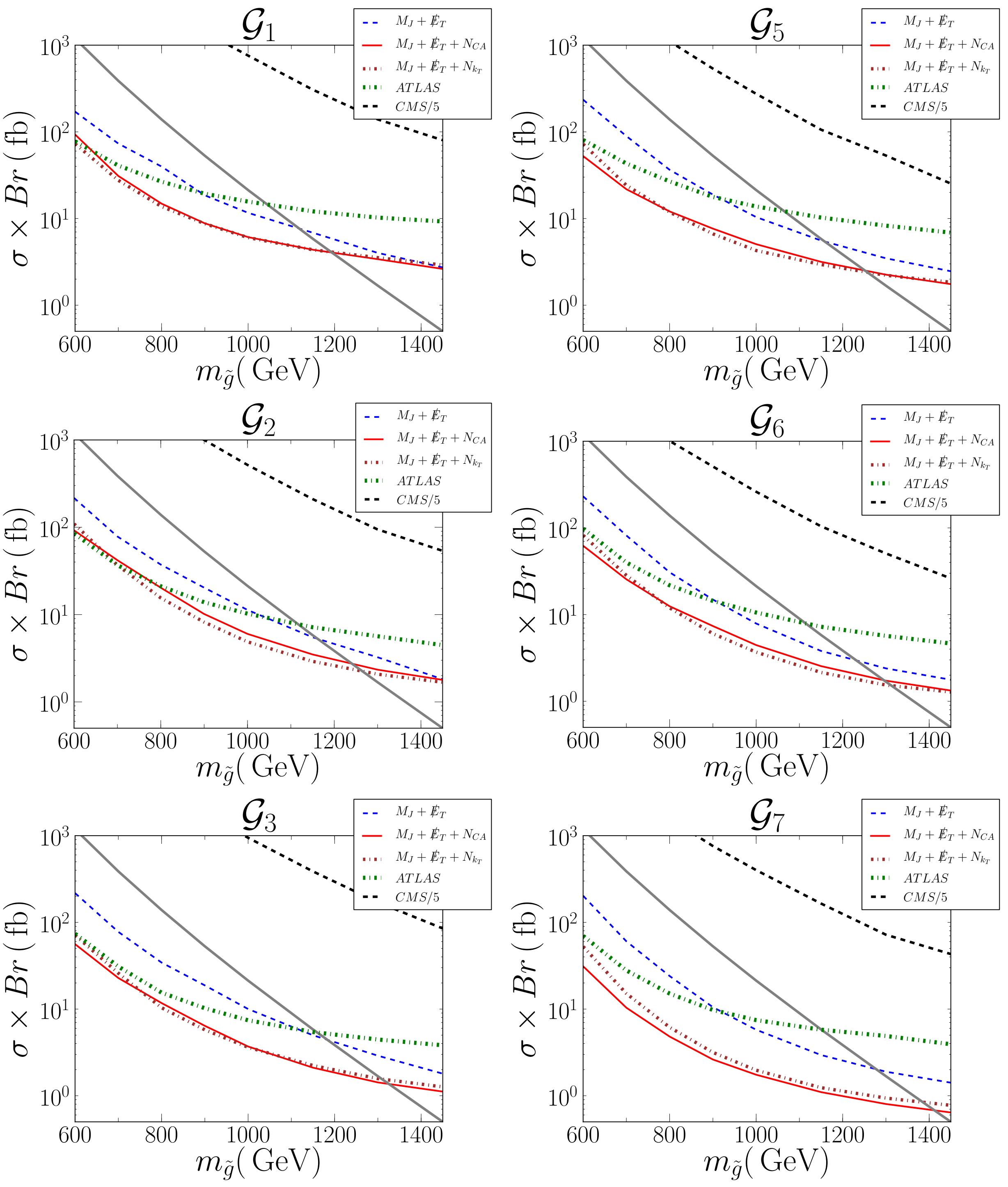

FIG. 12: 95% exclusion limits on $\sigma \times \mathcal{B}$ for the $M_J + \not\!\!E_T$ search (dashed blue), the $M_J{+}\not\!\!E_T{+}N_{\rm CA}$ search (solid red), and the $M_J{+}\not\!\!E_T{+}N_{\rm k_T}$ search (dash-dotted brown) for the R-parity conserving topologies $\mathcal{G}_1$, $\mathcal{G}_2$, and $\mathcal{G}_3$ (left, top to bottom) and the corresponding RPV ones, $\mathcal{G}_5$, $\mathcal{G}_6$, and $\mathcal{G}_7$ (right, top to bottom). The exclusion limits given by the ATLAS high multiplicity search [43] (dash-dotted green) and the CMS black hole search [44] (dashed black) as well as the NLO gluino production cross section (grey solid line) are also shown. The systematic uncertainty on the background is assumed to be 30%. Note that the CMS limit is rescaled by a factor of 5.

- $n_j \geq 7$ with $p_T \geq 80\,\mathrm{GeV}$
- $n_j \geq 8$ with $p_T \geq 80\,\mathrm{GeV}$

It is further required that events contain no isolated leptons and that $\not{E}_T/\sqrt{H_T} > 4\,\mathrm{GeV}^{\frac{1}{2}}$. Note that for an event with $H_T = 1000\,\mathrm{GeV}$ ($3000\,\mathrm{GeV}$) the latter cut corresponds to a $\not{E}_T$ cut of $126\,\mathrm{GeV}$ ($219\,\mathrm{GeV}$).

In order to compare the performance of the ATLAS search to the optimized search strategies in Tables 4 and 5, we assumed that the ATLAS search could be scaled up to $30\,\mathrm{fb}^{-1}$ while keeping the cuts fixed. The re-estimated expected limits have been computed by linearly rescaling the expected number of background events and the corresponding uncertainties (as given by ATLAS) to the new luminosity and computing the exclusion limit as outlined in Sec 5.3. Such a linear rescaling, which assumes that systematic uncertainties do not come to dominate the limits, is probably overly optimistic.

We find that for those benchmark signals with large final state multiplicities and intrinsic $\not{E}_T$ ($\mathcal{G}_{1,2,3}$ and $\mathcal{G}_{5,6,7}$), i.e. the sorts of signals that the ATLAS search is designed to be sensitive to, the M_J+$\not{E}_T$+N_{CA} search generally outperforms the ATLAS search, particularly at higher gluino masses, where the exclusion limit on $\sigma \times \mathcal{B}$ is improved by a factor $2-5$. For these benchmark signals this corresponds to an extended gluino mass reach of order $100\,\mathrm{GeV}$ to $250\,\mathrm{GeV}$. While it is difficult to know the extent to which this promising result can be realized at ATLAS or CMS, it is worth emphasizing that whatever the final search sensitivity should turn out to be, it is already valuable to have a search strategy that is governed by different systematic uncertainties.

5.5.2. CMS Black Hole Search

The CMS black hole search [44], which makes use of $4.7\,\mathrm{fb}^{-1}$ at 7 TeV, is also sensitive to high multiplicity final states. This search makes use of 16 search regions, each of which corresponds to different $S_{T\,\mathrm{min}}$ and N_{min} cuts. Here $S_{T\,\mathrm{min}} \in [1.9, 4.1]\,\mathrm{TeV}$ is a cut on the scalar sum over transverse energy and $N_{\mathrm{min}} \in [3, 7]$ is a cut on the total number of reconstructed objects with $E_T > 50$ GeV. See ref. [44] for details.

In order to compare the expected performance of the CMS search to the optimized search strategies in Tables 4 and 5, it is necessary to extrapolate the CMS background estimates from 7 TeV to 8 TeV. Because there are no missing energy requirements, the background is completely dominated by QCD events. The absence of any intrinsic high energy scale in the background allows us to adopt the following approximate extrapolation. For each value of N_{min}, CMS provides the expected number of background events as a function of S_T, which we fit to an exponential,

$$N_{\mathrm{QCD}}^{7\,\mathrm{TeV}}(S_T; N_{\mathrm{min}}) = e^{-\alpha(S_T - S_T^{(0)})} \tag{31}$$

where α and $S_T^{(0)}$ are fit parameters. The number of background events as a function of S_T at $8\,\mathrm{TeV}$ is then estimated to be

$$N_{\mathrm{QCD}}^{8\,\mathrm{TeV}}(S_T; N_{\mathrm{min}}) = N_{\mathrm{QCD}}^{7\,\mathrm{TeV}}\left(\frac{7\,\mathrm{TeV}}{8\,\mathrm{TeV}} \times S_T; N_{\mathrm{min}}\right) \tag{32}$$

These background estimates can then rescaled to $30\,\mathrm{fb}^{-1}$ and combined with the efficiencies of the benchmark signals in each of the 16 search regions to obtain the expected sensitivity of the CMS search at 8 TeV. While these background estimates are inexact, they are sufficient for demonstrating that the CMS search results in expected limits that are about two orders of magnitude weaker those obtained by a M_J+$\not{E}_T$+$N_{\rm CA}$ search (see Fig. 12). This is because for the gluino masses that are accessible at 8 TeV, the S_T cuts are highly inefficient, and the absence of missing energy requirements results in large QCD backgrounds. This demonstrates that high multiplicity searches targeting black holes are not necessarily well suited for other kinds of high multiplicity signals.

6. DISCUSSION

Recent years have seen an impressive amount of research on a large variety of jet substructure techniques.[7] The majority of this work has focused on the development of either general purpose tools (jet grooming, top tagging, etc.) or jet substructure analyses tailored to specific search channels (e.g. the BDRS boosted Higgs search [22]). One area that has seen less work is the design of search techniques for topologies that are more complicated or whose structure is not known a priori. In this paper we have taken a step in this direction by arguing that jet substructure suggests a different approach to counting jet multiplicities that results in an effective search strategy that is sensitive to a variety of high multiplicity topologies.

The flexibility inherent in this approach raises the possibility of loosening missing energy cuts in favor of well chosen jet substructure cuts. This is of special interest for new physics scenarios in which signals exhibit little or no intrinsic missing energy, such as supersymmetric scenarios with baryonic RPV. In Sec. 5.4 we have seen that for signals with large final state multiplicities and some (though not necessarily very much) intrinsic $\not{E}_T$, the introduction of $N_{\rm CA}$ cuts does in fact lead to lower $\not{E}_T$ requirements. While this represents only a modest push towards the regime of (near)-vanishing $\not{E}_T$ requirements, it is nevertheless an encouraging result given how effective $\not{E}_T$ requirements are in reducing the huge QCD backgrounds. In fact, we find that trading $\not{E}_T$ cuts for $N_{\rm CA}$ cuts is particularly effective for the QCD background—it is the need to suppress the $t\bar{t}$+jets background that prevents the $\not{E}_T$ cuts from being loosened further in Table 5. We anticipate that if additional handles were introduced to combat the $t\bar{t}$+jets background (e.g. vetoing on b-jets), then $\not{E}_T$ cuts could be loosened even further. We have not pursued this interesting direction here, since our goal was to keep the search strategy as inclusive as possible.

One possible concern with high multiplicity searches at the LHC is their potential sensitivity to pile-up, something that becomes more pressing as the LHC pushes towards higher and higher luminosities. In this paper we have advocated the use of jet trimming to reduce the present search's sensitivity to pile-up, but something like the technique introduced in ref. [46] might also be necessary. This whole issue would need to be revisited if the search were to be performed by ATLAS or CMS. This is particularly the case because it is impossible for us to thoroughly examine pile-up effects given their sensitivity to detector effects and the fact that each collaboration has detector specific methods for mitigating pile-up effects. It is worth pointing out that one possible advantage of $N_{\rm CA}$ is that it includes some built-in jet grooming by virtue of the veto it imposes on asymmetric subjet energy sharing.

In conclusion, we have seen that an effective search strategy can be developed by exploit-

[7] For a comprehensive set of references see the BOOST 2010 & 2011 proceedings [45].

ing missing energy, a sum over fat jet masses, and a sum over fat jet subjet counts. The two subjet counting algorithms presented, $N_{\rm CA}$ and $N_{\rm kT}$, yield comparable results, so that a choice between the two would need to be guided by experimental studies (with a particular focus on inherent systematic uncertainties, performance under pile-up, etc.). Other subjet counting algorithms are possible,[8] but what we would like to stress here is that, as has been seen in many other jet substructure studies, the flexibility of the fat jet approach is very powerful. In this case the potential for systematic data driven estimates of the QCD background is of particular importance. As another example of the flexibility of the fat jet approach, we refer the reader to the related work in ref. [47], which focuses on high multiplicity hadronic final states with vanishing missing energy.

Acknowledgements

We thank Timothy Cohen, Eder Izaguirre, Mariangela Lisanti, Jesse Thaler, Gavin Salam, Stefan Höche, Steffen Schumann, Ariel Schwartzman, Ken Van Tilburg, and Xinlu Huang for helpful discussions. MJ would like to thank Michael Spannowsky for interesting conversations. Special thanks to Timothy Cohen for having provided Monte Carlo data for the QCD background. SE and JW are supported by the US DOE under contract number DE-AC02-76-SF00515. SE is supported by a Stanford Graduate Fellowship. AH is supported by the US DOE under contract number DE-FG02-90ER40542.

Appendix A: Genetic algorithm for optimizing search regions

The genetic algorithm is initialized with 1000 search strategies, where each search strategy is a set of search regions. Each of these search strategies is formed as follows. First a random selection of 40 of the $N_{\rm cuts}$ search regions is chosen. Each of these 40 search regions is assigned a weight proportional to the number of models it covers with $\mathcal{E} \leq \mathcal{E}_{\rm crit}$. Finally, 1000 search strategies are created by sampling (without replacement) from these 40 search regions. This gives a slight preference in the initialization stage to search regions that are sensitive to more models. The exact initialization procedure is not critical for rapid convergence of the algorithm

The search strategies are evaluated to see how many models they cover within the desired efficacy, and a "fitness" is assigned to them with the formula

$$f(C,M) = \frac{1}{M_{\rm max}^2 - (M^2 - C)}\,, \tag{A1}$$

where M is the number of models covered, C is the number of search regions in the search strategy, and $M_{\rm max}$ is the total number of models. This fitness function strongly penalizes search strategies that do not cover all models, followed by a penalty for having too many search regions.

[8] We have investigated the possibility of a subjet counting algorithm based on N-subjettiness [19], using a boosted decision tree to map τ_N space to different subjet multiplicities. Although the resulting algorithm performed worse than $N_{\rm CA}$ and $N_{\rm kT}$, it is possible that a more thorough study could lead to improvements.

After evaluating the fitness of the search strategies, the least fit 50% are removed. Pairs of fit search strategies are then selected and a new search strategy is created by taking a randomly determined fraction of each search strategy's search regions. For instance, if the two selected search strategies had N_1 and N_2 search regions, then a uniform random number on the unit line segment, x, would determine that xN_1 search regions would be taken from the first search strategy and $(1-x)N_2$ would be taken from the second search strategy. So if $N_1 = 20$ and $N_2 = 30$ and $x = 0.20$, 4 search regions would be taken from the first search strategy and 24 would be taken from the second. If duplicate signal regions are selected, the duplicate is removed, reducing the number of search regions. After creating a new search strategy, the search is mutated to guarantee that the population of search strategies has sufficient diversity. Each search region within a search strategy has a finite probability of being changed to another random search region. We use 6% for this probability known as the "mutation rate". Thus for the 16 search regions in the example, 1 change would be made on average.

If after ten consecutive generations no progress has been made, i.e. if no solution has been found that covers the entire model space, then a solution is manually created by forcing every model to be covered by some search region. This can be done by increasing the number of search regions in the search strategies until full coverage is achieved. Finally, if every model is covered and no further progress is achieved for seven generations, search strategies are scoured to see if any search regions can be removed without reducing coverage. Either way, the genetic algorithm is restarted. If no progress in reducing the number of search regions in a search strategy has been made in twenty generations, the program ends.

Typically, the algorithm converges after 20 to 30 generations, and 10 to 30 distinct optimized search strategies are found each time. While the termination of the program does not guarantee that the optimal solution has been found, re-running the program multiple times usually results in the same number of required search regions. The resulting search strategies typically have similar features even if they differ slightly in detail.

[1] S. Dimopoulos and G. L. Landsberg, Phys. Rev. Lett. **87**, 161602 (2001) [hep-ph/0106295].

[2] J. Bramante, J. Kumar and B. Thomas, Phys. Rev. D **86**, 015014 (2012) [arXiv:1109.6014 [hep-ph]];

[3] M. Papucci, J. T. Ruderman and A. Weiler, JHEP **1209**, 035 (2012) [arXiv:1110.6926 [hep-ph]].

[4] J. Kumar, arXiv:1211.6503 [hep-ph]. H. Baer, V. Barger, P. Huang and X. Tata, JHEP **1205**, 109 (2012) [arXiv:1203.5539 [hep-ph]]. T. Li, J. A. Maxin, D. V. Nanopoulos and J. W. Walker, arXiv:1108.5169 [hep-ph]. G. L. Kane, E. Kuflik, R. Lu and L. -T. Wang, Phys. Rev. D **84**, 095004 (2011) [arXiv:1101.1963 [hep-ph]]. D. Feldman, G. Kane, E. Kuflik and R. Lu, Phys. Lett. B **704**, 56 (2011) [arXiv:1105.3765 [hep-ph]]. B. S. Acharya, P. Grajek, G. L. Kane, E. Kuflik, K. Suruliz and L. -T. Wang, arXiv:0901.3367 [hep-ph].

[5] J. T. Ruderman, T. R. Slatyer and N. Weiner, arXiv:1207.5787 [hep-ph].

[6] B. Bhattacherjee, J. L. Evans, M. Ibe, S. Matsumoto and T. T. Yanagida, arXiv:1301.2336 [hep-ph].

[7] D. Curtin, R. Essig and B. Shuve, arXiv:1210.5523 [hep-ph].

[8] C. T. Hill, Phys. Lett. B **266**, 419 (1991).

[9] D. A. Dicus, B. Dutta and S. Nandi, Phys. Rev. D **51**, 6085 (1995) [hep-ph/9412370].

[10] C. Gross, G. M. Tavares, C. Spethmann and M. Schmaltz, arXiv:1209.6375 [hep-ph].

[11] E. Gerwick, T. Plehn, S. Schumann and P. Schichtel, JHEP **1210**, 162 (2012) [arXiv:1208.3676 [hep-ph]].

[12] E. Gerwick, B. Gripaios, S. Schumann and B. Webber, arXiv:1212.5235 [hep-ph].

[13] A. Hook, E. Izaguirre, M. Lisanti and J. G. Wacker, Phys. Rev. D **85**, 055029 (2012) [arXiv:1202.0558 [hep-ph]].

[14] S. Catani, Y. L. Dokshitzer, M. H. Seymour and B. R. Webber, Nucl. Phys. B **406**, 187 (1993).

[15] M. Wobisch and T. Wengler, In *Hamburg 1998/1999, Monte Carlo generators for HERA physics* 270-279 [hep-ph/9907280].

[16] M. Cacciari, G. P. Salam and G. Soyez, Eur. Phys. J. C **72**, 1896 (2012) [arXiv:1111.6097 [hep-ph]]; M. Cacciari and G. P. Salam, Phys. Lett. B **641**, 57 (2006) [hep-ph/0512210].

[17] M. Cacciari, G. P. Salam and G. Soyez, JHEP **0804**, 063 (2008) [arXiv:0802.1189 [hep-ph]].

[18] D. Krohn, J. Thaler and L. -T. Wang, JHEP **1002**, 084 (2010) [arXiv:0912.1342 [hep-ph]].

[19] J. Thaler and K. Van Tilburg, JHEP **1103**, 015 (2011) [arXiv:1011.2268 [hep-ph]]. J. Thaler and K. Van Tilburg, JHEP **1202**, 093 (2012) [arXiv:1108.2701 [hep-ph]].

[20] G. Aad *et al.* [ATLAS Collaboration], JHEP **1205**, 128 (2012) [arXiv:1203.4606 [hep-ex]].

[21] See e.g. G. P. Salam, Eur. Phys. J. C **67**, 637 (2010) [arXiv:0906.1833 [hep-ph]].

[22] J. M. Butterworth, A. R. Davison, M. Rubin and G. P. Salam, AIP Conf. Proc. **1078**, 189 (2009) [`arXiv:0809.2530`]. J. M. Butterworth, A. R. Davison, M. Rubin and G. P. Salam, Phys. Rev. Lett. **100**, 242001 (2008) [`arXiv:0802.2470`].

[23] T. Plehn, G. P. Salam and M. Spannowsky, Phys. Rev. Lett. **104** (2010) 111801 [arXiv:0910.5472 [hep-ph]].

[24] T. Plehn, M. Spannowsky, M. Takeuchi and D. Zerwas, JHEP **1010**, 078 (2010) [`arXiv:1006.2833`].

[25] T. Stelzer and W. F. Long, "Automatic generation of tree level helicity amplitudes", Comput. Phys. Commun. 81 (1994) 357371, [hep-ph/9401258].

[26] F. Maltoni and T. Stelzer, "MadEvent: Automatic event generation with MadGraph", JHEP 02 (2003) 027, [hep-ph/0208156].

[27] J. Alwall, P. Demin, S. de Visscher, R. Frederix, M. Herquet, F. Maltoni, T. Plehn and D. L. Rainwater *et al.*, JHEP **0709**, 028 (2007) [arXiv:0706.2334 [hep-ph]].

[28] T. Sjostrand, S. Mrenna and P. Z. Skands, JHEP **0605**, 026 (2006) [hep-ph/0603175].

[29] M. Mangano (2002), Fermilab Monte Carlo Workshop, Oct. 2002 (unpublished).

[30] S. Alekhin, J. Blumlein, S. Klein and S. Moch, Phys. Rev. D **81**, 014032 (2010) [arXiv:0908.2766 [hep-ph]].

[31] T. Gleisberg, S. Hoeche, F. Krauss, M. Schonherr, S. Schumann, et al., "Event generation with SHERPA 1.1", JHEP 0902 (2009) 007, arXiv:0811.4622 [hep-ph].

[32] F. Krauss, R. Kuhn, and G. Soff, "AMEGIC++ 1.0: A Matrix element generator in C++", JHEP 0202 (2002) 044, arXiv:hep-ph/0109036 [hep-ph].

[33] S. Schumann and F. Krauss, "A Parton shower algorithm based on Catani-Seymour dipole factorisation", JHEP 0803 (2008) 038, arXiv:0709.1027 [hep-ph].

[34] T. Gleisberg and S. Hoeche, "Comix, a new matrix element generator", JHEP 0812 (2008) 039, arXiv:0808.3674 [hep-ph].

[35] S. Hoeche, F. Krauss, S. Schumann, and F. Siegert, "QCD matrix elements and truncated showers", JHEP 0905 (2009) 053, arXiv:0903.1219 [hep-ph].

[36] S. Catani, F. Krauss, R. Kuhn and B. R. Webber, JHEP **0111**, 063 (2001) [hep-ph/0109231].

[37] J. Conway, A Pretty Good Simulation (2009),

http://physics.ucdavis.edu/ conway/research/ software/pgs/pgs4-general.htm.

[38] [ATLAS Collaboration], ATLAS-CONF-2012-109.

[39] W. Beenakker, R. Hopker and M. Spira, hep-ph/9611232; W. Beenakker, R. Hopker, M. Spira and P. M. Zerwas, Nucl. Phys. B **492**, 51 (1997) [hep-ph/9610490].

[40] [ATLAS Collaboration], ATLAS-CONF-2012-065.

[41] D. S. M. Alves, E. Izaguirre and J. G. Wacker, JHEP **1110**, 012 (2011) [arXiv:1102.5338 [hep-ph]].

[42] R. Essig, E. Izaguirre, J. Kaplan and J. G. Wacker, JHEP **1201**, 074 (2012) [arXiv:1110.6443 [hep-ph]].

[43] [ATLAS Collaboration], ATLAS-CONF-2012-103.

[44] S. Chatrchyan *et al.* [CMS Collaboration], JHEP **1204**, 061 (2012) [arXiv:1202.6396 [hep-ex]].

[45] A. Abdesselam, E. B. Kuutmann, U. Bitenc, G. Brooijmans, J. Butterworth, P. Bruckman de Renstrom, D. Buarque Franzosi and R. Buckingham *et al.*, Eur. Phys. J. C **71**, 1661 (2011) [arXiv:1012.5412]; A. Altheimer, S. Arora, L. Asquith, G. Brooijmans, J. Butterworth, M. Campanelli, B. Chapleau and A. E. Cholakian *et al.*, [arXiv:1201.0008].

[46] G. Soyez, G. P. Salam, J. Kim, S. Dutta and M. Cacciari, arXiv:1211.2811 [hep-ph].

[47] T. Cohen, E. Izaguirre, M. Lisanti and H. K. Lou, arXiv:1212.1456 [hep-ph].